

## Portland State University PDXScholar

---

Dissertations and Theses

Dissertations and Theses

---

Fall 12-13-2013

# An Investigation into Instructional Support for Data Analysis in High School Science Inquiry

Anika Rae Baker-Lawrence  
*Portland State University*

Let us know how access to this document benefits you.

Follow this and additional works at: [http://pdxscholar.library.pdx.edu/open\\_access\\_etds](http://pdxscholar.library.pdx.edu/open_access_etds)

 Part of the [Educational Methods Commons](#), [Science and Mathematics Education Commons](#),  
and the [Secondary Education and Teaching Commons](#)

---

### Recommended Citation

Baker-Lawrence, Anika Rae, "An Investigation into Instructional Support for Data Analysis in High School Science Inquiry" (2013).  
*Dissertations and Theses*. Paper 1534.

[10.15760/etd.1534](https://doi.org/10.15760/etd.1534)

This Thesis is brought to you for free and open access. It has been accepted for inclusion in Dissertations and Theses by an authorized administrator of PDXScholar. For more information, please contact [pdxscholar@pdx.edu](mailto:pdxscholar@pdx.edu).

An Investigation into Instructional Support for  
Data Analysis in High School Science Inquiry

by

Anika Rae Baker-Lawrence

A thesis submitted in partial fulfillment of the  
requirements for a degree of

Master of Science in Teaching  
in  
General Science

Thesis Committee:  
William Becker, Chair  
Emily Saxton  
Melissa Potter

Portland State University  
2013

© 2013 Anika Rae Baker-Lawrence

## Abstract

The implementation of scientific inquiry in the high school classroom has proven to be not only relevant and exploratory, but challenging and engaging as well. This style of curriculum design has been recognized as a primary means of achieving the goals and objectives set by the National Resource Council (NRC, 1996). While much research has shown that science inquiry helps students to gain understanding of content knowledge, little research has been conducted to assess gains in higher order thinking skills, specifically those related to data analysis (Anderson, 2002; Germann and Aram, 1996; Hofstein, Navon, Kipnis, and Mamlok-Naaman, 2005; Miner, Levy, and Century, 2009; Windschitl, Thompson, and Braaten, 2008; Zohar and Dori, 2003). Through a better understanding of the scientific inquiry process as well as insights into students' struggles with data analysis, we can better understand how to effectively implement strategies in the classroom that encourage the higher order thinking skill of data analysis. This mixed methods, multiple-case study investigated teacher practice in eight high school science inquiry units in the Portland, Oregon metropolitan area and the data analyses that students produced in their accompanying work samples. The results of this study indicate that students struggle to produce proficient analysis and interpretations of data. The areas of student struggle were in the areas that required higher order thinking: analyzing results, drawing conclusions, and communicating results. Furthermore, this research discusses areas of data analysis instruction that may benefit from professional development opportunities.

## Table of Contents

Abstract .....	i
Introduction .....	1
Literature Review .....	5
Scientific Inquiry for Higher Order Cognition .....	5
Student Struggles with Data Analysis .....	10
Suggestions for Instructional Practices that Encourage Data Analysis Skills .....	13
Summary .....	18
Research Question .....	20
Methods .....	22
Overview .....	22
Participants .....	23
Data Collection Instruments .....	25
Procedure .....	27
Results .....	34
Case Descriptions .....	34
Abigail.....	34
Cathy .....	35
Felicia .....	36
Joe .....	37
Margot .....	38

Penny .....	40
Sonia .....	41
Teacher Instructional Portfolio Results .....	42
Quantitative TIP Results.....	42
Qualitative TIP Results.....	43
Student Work Sample Results.....	46
Quantitative SWS Results.....	46
Qualitative SWS Results.....	47
Comparative Results .....	48
Comparative Quantitative Results. ....	48
Comparative Qualitative Results .....	50
Discussion .....	51
Trends in Teacher Instruction .....	51
Classroom Roles. ....	52
Content and Cognitive Skills .....	54
Assessment for Learning .....	55
General Instructional Practices .....	58
Data Analysis Specific Instructional Practices .....	59
Data Analysis Best Practices .....	61
Trends in Teacher Instruction Summary .....	62
Trends in Student Work Samples .....	65
Quantitative Trends in SWS.....	65

Qualitative Trends in SWS.....	66
Trends in Students' Data Analysis Skills Summary .....	75
Case Study Comparisons .....	78
Low TIP – Low SWS – Cathy, Margot, and Penny .....	78
High TIP – Low SWS – Felicia and Sonia .....	80
High TIP – High SWS – Abigail and Sonia .....	82
Case Study Comparison Summary .....	83
Discussion Summary .....	84
Limitations of the Study .....	88
Questions for Further Research .....	91
Implications for Instructional Practice .....	93
References .....	96
Appendix A .....	100
Appendix B .....	105
Appendix C .....	107
Appendix D .....	109
Appendix E .....	110

## Introduction

To secure future prosperity, we need to ensure that our students have the “21<sup>st</sup> century skills” they will need to be successful. Our current system of education and assessment emphasizes rote memorization above a more complete and well-rounded understanding that accompanies higher order cognition (Elby, 1999; Linn et al., 2006; Schoenfield, 1988; Stern and Ahlgren, 2002). In contrast, 21<sup>st</sup> century skills prioritize higher order thinking and problem-solving combined with relevant content knowledge (Silva, 2009). “Higher order thinking” refers to the critical, logical, or metacognitive skills students apply when they encounter novel problems, questions, or situations (King, Goodson, and Rohani, 1998). These skills are in contrast to “lower order thinking” skills, which are based in retention of simple, more certain, factual information (NRC, 1987). With specific regard to science, students’ successful application of higher order thinking skills results in rich data analyses and scientific explanations, and further promotes the continued growth of these and other intellectual skills across disciplines.

To combat the trend of teaching to lower-order thinking, the Next Generation Science Standards have adopted scientific inquiry as a means of increasing students’ conceptual understanding and scientific skill set (National Research Council, 2012). Many studies have shown that utilizing scientific inquiry in the classroom can produce gains in learning (Anderson, 2002; Germann and Aram, 1996; Hofstein, Navon, Kipnis, and Mamlok-Naaman, 2005; Miner, Levy, and Century, 2009; Windschitl, Thompson,



and Braaten, 2008; Zohar and Dori, 2003). Scientific inquiry has the potential to be an iterative, cyclical process that can allow the participant to engage in multiple experiences with higher order thinking (Anderson, 2002; Germann and Aram, 1996; Miner, Levy, and Century, 2009). Among these higher-order thinking skills is the ability to understand and analyze data.

Data analysis is a multifaceted process in which raw data are transformed into useful information that informs conclusions and ultimately supports decision-making. Data analysis has proven to be one of the skills that students struggle with throughout their academic and professional careers (Hug and McNeill, 2008; Germann and Aram, 1996; Lovett and Chang, 2007). Thus, ensuring that educators have the means to properly teach and assess their students' ability to analyze data is a worthy endeavor. Furthermore, as many states begin to adopt the Next Generation Science Standards, teachers will need to continue participating in professional development trainings on how to effectively implement scientific inquiry in their classrooms. Thus, it is with necessity that the study of instructional techniques related to data analysis in science inquiry be investigated.

Although the Next Generation Science Standards encourage the implementation of science inquiry instruction in K-12 classrooms, the standards do not dictate a single approach (NRC, 2012). Instead, the NRC (2012) advises teachers to use a variety of instructional practices to support knowledge development, understandings, and skills. A sampling of instructional methods identified in the research literature for teaching data analysis skills includes the use of explicit instruction (Keiler, 2007; Lovett and Chang,

2007), scaffolding of assignments (Germann and Aram, 1996), and incorporation of formative assessments (Ruiz-Primo and Furtak, 2008). Other science-specific techniques that have been investigated include the use of first and second hand data sources (Hug and McNeill, 2008), use of both covariant and non-covariant data types (Kanari and Millar, 2004), and the integration of math and science (Keiler, 2007).

While this and other research has investigated different methods of developing data analysis skills in the K-12 classroom, the literature has primarily assessed gains in learning via gains in students' content knowledge, with little to no assessment of students' thinking or reasoning skills (Germann and Aram, 1996; Hofstein, Navon, Kipnis, and Mamlok-Naaman, 2005; Miner, Levy, and Century, 2009; Windschitl, Thompson, and Braaten, 2008; Zohar and Dori, 2003). Content knowledge is indeed important, many would argue particularly so in the current "teach to the test" climate (Higgins, Miller, and Wegmann, 2006; Lomax et al., 1995; Wiggins, 2001). However, this study investigates best practices for teaching and assessing students' gains in the higher order cognitive skills of data analysis.

The purpose of this mixed-methods, multiple-case study is to investigate the effects of teacher instructional practices on high school students' ability to analyze science inquiry data. The research questions are threefold: first, what instructional practices were implemented as documented by the Teacher Instructional Portfolio (TIP)? Second, what are the trends in students' ability to analyze data in their Student Work Samples (SWS)? Finally, is there a relationship between instructional practices and students' ability to analyze data? The TIPs and SWSs from eight high school science

inquiry units were analyzed. Both quantitative and qualitative analyses were conducted utilizing rubrics and researcher-developed coding schemes to characterize teacher instruction and student performance around data analysis, and to investigate whether any relationships exist between the teacher instructional practices and the students' ability to analyze data.

## Literature Review

Proper implementation of science inquiry in the high school classroom can produce positive gains in content knowledge, but do students gain the higher order thinking skills required for data analysis? Through a better understanding of the scientific inquiry process as well as insights into students' struggles with data analysis, we can more effectively implement strategies in the classroom that support students in developing the higher order thinking skills required for data analysis.

### **Scientific Inquiry for Higher Order Cognition**

In order for students to have an authentic science experience in the classroom, Windschitl, Thompson, and Braaten (2008) assert that genuine forms of inquiry should model the qualities of "testability, conjecture, explanation, principled revision, and generativity." Utilizing multiple data sources in the investigation of more than 60 cases of pre-service teachers, Windschitl et al. (2008) found that scientific inquiry moves the classroom learning process away from the most commonly practiced form of investigation in our classrooms- the scientific method. The authors suggest that instead of the scientific method, teachers can implement science inquiry investigations in their classrooms to effectively increase science concept understanding. Scientific inquiry is defined by the authors as an intentionally loosely-constructed system made up of four key components: 1) organization of what is known and what is to be learned 2) generation of a testable hypothesis, 3) gathering of evidence, and 4) construction of an

argument. Though this may appear at first glance to be another linear process, the author asserts that this is an “organic process” that allows for learners as scientists to constantly reexamine multiple steps along the way (Windschitl, Thompson, and Braaten, 2008). Through science inquiry, students can experience science as an investigation that not only seeks to deepen their understanding of science but encourages higher order thinking and questioning as well.

Science inquiry has the ability to help encourage higher order thinking, but there is not just one form of science inquiry. In 2005, Hofstein, Navon, Kipnis, and Mamlok-Naaman investigated students’ ability to ask higher order questions in closed and open-ended inquiry style chemistry laboratories. This two-part study first investigated high schools students’ ability to ask questions in general chemistry inquiry experiences. The second part of this investigation examined how high school students could ask questions in a novel learning context, specifically during the analysis of a scientific article. Working over the course of five years in Israel, nearly 100 general chemistry inquiry style labs were created for 11<sup>th</sup> and 12<sup>th</sup> grade classrooms ranging in topic from acid-base reactions and stoichiometry to equilibrium and rates of reactions. This study specifically examined the results of six 12<sup>th</sup> grade chemistry classes over the course of 2 years, with 111 students ranging in age from 17-18. The students were split into two groups, the “inquiry group” (N=55) and the “traditional group” (N=56). Split into their research groups, the students were provided with similar resources: the same syllabi, the same textbooks, and the same grouping format (small groups of 3-4 students). The traditional group followed only scripted instructions from a laboratory manual during a process

called “close-ended” or confirmatory, inquiry. The inquiry group first conducted closed-ended inquiry and then “open-ended” inquiries, in which students were asked to design and conduct their own chemistry inquiry experiment based off the results of their close-ended laboratory results. Near the end of their 2-year chemistry education, having conducted approximately 15 laboratories, the students were asked to complete two tasks. The first task was for students to conduct a simple inquiry experiment (mixing two chemicals) and record all relevant questions the investigation prompted in their thoughts. The second task asked students to closely read a scientific article and, again, record all of relevant questions the article prompted. Hofstein et al. (2005) found that, with regard to both tasks, the high school students who participated in inquiry groups were able to ask higher-order questions. These were not surprising results, in that the students who participated in the inquiry group had multiple opportunities to ask higher order questions during their open-ended experiences. This research suggests that science classroom experiences should include more open-ended inquiries to promote students’ ability to ask higher order questions.

In addition to helping students ask higher-order questions, scientific inquiry has the potential to help students develop other necessary higher order thinking skills. Miner, Levy, and Century (2009) carefully selected, reviewed, coded, and analyzed 138 studies conducted between 1984 and 2002 with K-12 students participating in inquiry-based instruction. Utilizing first descriptive statistics and later more complex techniques, including a high rigor MANOVA, the authors concluded that students who received science inquiry instruction had improved learning outcomes when compared

with their content knowledge understanding prior to instruction, as well as with students who received a different type of instruction that was less inquiry-focused. Specifically, the authors found that there was a statistically significant improvement in the students' higher-order cognition associated with participation in science inquiry, particularly when linked with an opportunity to process the meaning of their experimental findings through class discussions. The authors attribute these learning gains to a combination of instructional elements associated with science inquiry. The authors described the results of effectively implemented inquiry as an emphasis of student 1) self-efficacy, 2) active thinking, and 3) motivation. Each of these elements is essential in allowing students to gain and demonstrate their ability to access knowledge at a higher level.

A general consensus among educators is that students should leave the classroom prepared to answer the questions that the future has to offer. However, the educational community has also generally agreed that there is a gap between students' levels of achievement and their higher-level cognitive abilities. In an effort to investigate this achievement gap, Zohar and Dori (2003) specifically investigated higher and lower achieving students' learning gains with regard to higher order thinking skills in the context of science classrooms. Structuring their research within the framework of the "Science Technology Society" approach, which works to make science meaningful to all students through the emphasis of higher order thinking and problem solving, the researchers undertook a four-part study. Study 1 examined seven classes of 10<sup>th</sup> grade students' question-posing capabilities as fostered through a case-based teaching and

learning module on real world controversial topics related to air quality. Study 2 investigated two classes of middle school students' argumentation skills as fostered through a case-based teaching and learning module on bioethical topics related to genetics. Study 3 investigated eight classes of 10<sup>th</sup> to 12<sup>th</sup> grade students' higher order thinking skills as fostered through case studies in biotechnology. Study 4 examined 21 classrooms of seventh grade students' critical thinking abilities and scientific thinking as assessed by two open response tests. After disaggregating and analyzing the results of the four studies, Zohar and Dori (2003) found that in every study both higher and lower achieving students had positive gains in learning. Interestingly, Study 3 investigating higher order thinking was the only section of this research study where lower achieving students in fact had higher gains in learning than their high achieving counterparts, thus helping to reduce the achievement gap. The results of this research project suggest that for both high and low achieving students to be successful, higher order cognitive skills need to be emphasized alongside content knowledge, and that one should not be cast aside at the expense of another. Furthermore, teachers must be considerate of all students' abilities with regard to higher order thinking and problem solving in the context of science and technology, whether a student is high or low achieving.

Science inquiry provides students with the opportunity to not only develop their conceptual understandings of science, also to encourage the development of their higher order thinking and questioning skills (Windschitl, Thompson, and Braaten, 2008). When students are provided with multiple opportunities to experience the iterative science inquiry cycle, they are more prepared to ask and answer the deep, higher order



thinking questions necessary to be an effective problem solver in life (Hofstein et al., 2005). Furthermore, effectively implemented science inquiry has potential to activate student thinking and motivation, thus promoting students' self-efficacy (Miner, Levy, and Century, 2009). All students can benefit from and should therefore have access to understanding at a higher level, whether they are high or low achieving students (Zohar and Dori, 2003). As such, the science inquiry process is an effective technique for encouraging higher order cognition in our classrooms.

### **Students' Struggles with Data Analysis**

Assuming that science inquiry can be an effective tool for promoting students' higher order cognitive skills, it is important to all examine how students are currently able to participate in said inquiries. In a research study undertaken to inform the development of the National Committee on Science Education Standards and Assessment's (1994 draft) rubrics to assess 7<sup>th</sup> grade students' ability to perform science inquiry processes, Germann and Aram (1996) investigated the "processes of recording data, analyzing data, drawing conclusions, and providing evidence." Using an assessment of students' science process skills, the responses of 364 students were considered in the creation of the rubric. Germann and Aram (1996) identified three common areas where students often encounter difficulties while attempting to manipulate data skills sets: 1) analyzing results, 2) drawing conclusions, and 3) argument construction. These higher-level data analysis skills ask students to not just collect information, but instead to make sense of evidence based in scientific reasoning.

Hug and McNeill (2008) found similar results during their investigation into the influence of data on conversations in the classroom. Conducting their study in two urban middle school classrooms in a large Mid-western U.S. city, the researchers employed two curriculum units, one 7<sup>th</sup> grade unit on Chemistry (“How can I make new stuff from old stuff?”) and one 8<sup>th</sup> grade Biology unit (“Who will survive?”). Utilizing a scaffolded approach in developing scientific inquiry labs, the researchers paid close attention to students’ skills with specific attention to the sections on data analysis interpretation. Studying three chemistry and biology lessons each, the researchers analyzed classroom video footage against a coding scheme they developed to focus on common student difficulties surrounding multiple higher order thinking skills, including data analysis. The researchers detail three common difficulties that students encounter when analyzing data: 1) determining what kind of data can be used as evidence, 2) identifying accurate patterns from data, and 3) drawing and justifying conclusions from evidence. Each of these difficulties is an example of students’ struggling with higher order thinking, and if educators can appropriately identify where their students are struggling they can more effectively choose what teaching strategies will help to alleviate gaps in comprehension and thinking skill deficits.

In pursuit of an efficient integration of both the theoretical and applied perspectives of student learning, Lovett and Chang (2007) undertook a multipart research study that critically examined how university undergraduate students learn to analyze data. Working with students mostly from Carnegie Mellon’s College of Humanities and Social Science, the researchers utilized an assortment of approaches to

examine how students analyze data. Lovett and Chang (2007) found that, compared with expert participants, the novice undergraduates failed to plan how they were going to analyze their data. The authors exposed how students' lack of preparatory skills affects their ability to select appropriate analysis tools. Furthermore, the researchers found that students were inclined to choose quicker, more superficial approaches to data analysis, rather than using the more advanced manipulations to execute these analyses. The researchers go so far as to suppose that students have a "perverse efficiency" in learning what process steps are useful for successful problem solving during classroom-based inquiry activities, but often fail to master more general techniques used by scientists in the field (Lovett and Chang, 2007). It is clear that in order for students to appreciate the transferrable skills that data analysis has to offer, the value of these skills must be explicitly stated and emphasized throughout the classroom data analysis experience.

Germann and Aram (1996), Hug and McNeill (2008), and Lovett and Chang (2007) each found that the majority of students in their studies experienced struggles in their abilities to effectively analyze their data. Germann and Aram (1996) highlighted that the greatest areas of student struggle were in the skills of analyzing results, drawing conclusions, and constructing an argument. Hug and McNeill (2008) found very similar results as Germann and Aram (1996), describing the main areas of student struggle with data analysis as determining what kind of data can be used as evidence, identifying accurate patterns from data, and drawing and justifying conclusions from evidence. Furthermore, Lovett and Chang (2007) found that students had an impressive ability to

isolate content from cognition, in that they could perform calculations with semi-expert skill but were unable to explain the significance of their data analyses. In order for students to draw meaningful insights from their investigations, they need to be able to properly analyze their data. If students are unable to do so, it is worthwhile to investigate tasks that can encourage students to successfully analyze data.

### **Suggestions for Instructional Practices that Encourage Data Analysis Skills**

As part of a multipart investigation into the data analysis abilities of high school students, Keiler (2007) investigated the transferability of data handling. Analyzing classroom observations, student and teacher interviews, and the work samples of 60 students from 11 classrooms at a school chosen for its “typicality”; Keiler identified that “transfer events” were often the only source of students’ understandings of data analysis. The author found strong evidence that data analysis skills students exhibited in their science coursework were a direct result of horizontal transfer from their mathematics coursework. When the authors further questioned students to learn where data analysis was utilized in their mathematics courses, they found evidence that no single moment of learning had taught students their abilities but rather the culmination of years of vertical transfer. This transfer was further divided into “high road” (as in transfer that was deliberately applied to novel situations) and “low road” (as in transfer that was automatically applied to novel situations), as well as “positive transfer” (transfer that increased student comprehension in novel situations) and “negative transfer” (transfer that caused student resistance to comprehension in novel

situations). Finally, the author noted that to high performingly utilize students' ability to transfer data analysis skills from one set of coursework to another, instruction must be tailored to encourage transferability. Ultimately, the author found that if students were not specifically requested to employ their data analysis abilities, they would often neglect these skills in their work samples.

In addition to identifying student struggles with data analysis, Germann and Aram (1996) investigated what instructional practices are necessary for students to gain these essential skills. The authors concluded that students need high performing scaffolding to conduct investigations that will help them effectively understand the cause and effect nature of scientific investigations (Germann and Aram, 1996). The authors assert that a well-designed and scaffolded science inquiry assessment instrument should provide students with multiple opportunities to demonstrate their data analysis abilities. Such an assessment tool should include the skill sets of 1) induction, 2) identification of variables, 3) designing a data table, 4) experimentation and data recording, 5) analysis of results, 6) ability to draw conclusions, and 7) aptitude at providing evidence for said conclusions. From this research, it is clear that scaffolding science inquiries provide students with the opportunity to gain scientific understandings, but deliberately designed assessments must be created to intentionally assess student comprehension of both content and higher order cognition.

In their study investigating the influence of data on conversations as discussed above, Hug and McNeill (2008) also explored instructional strategies to help promote student success in data analysis. Hug and McNeill (2008) specifically explored whether

utilizing first-hand or second-hand data types were more effective in the middle school classroom. The authors discovered that during first-hand data experiences, the students had greater focus on “gathering data, discussing the limitations of the data, and identifying the data sources.” Conversely, during second-hand data experiences, students were more likely to “engage in data manipulation, identification of patterns, draw conclusions from data, and consider content knowledge” (Hug and McNeill, 2008). For students to properly analyze data they need experience with multiple data sources, because first-hand and second-hand data offer students different opportunities to engage in higher order cognition.

In another study focusing on data analysis, Kanari and Millar (2004) investigated how students draw conclusions from data collection and interpretations. The researchers conducted a two-phase investigation with a sample of 60 middle and high school students (20 students age 10, 20 students age 12, and 20 students age 14). In the first phase, each student was asked to collect and interpret data from an experiment involving variables that do have an effect on one another (“covariant”). In the second phase of the investigation, the same set of students was asked to collect and interpret data from a similar experiment, this time involving variables that do not have an effect on one another (“non-covariant”). To control variation that arises from hypothesis prediction, the authors provided the students with three pre-formulated hypotheses to choose from, which the students could exchange throughout the procedure if they changed their minds during their observations. Video recordings of the students’ experiments and audio recordings of student interviews were analyzed for both

investigations. The authors found no differences resulting across contexts or age groups, but did find substantially significant differences between the covariant and non-covariant investigations. Whereas all the students investigating covariant experiments were able to reach a correct conclusion, only 50% of the students reached correct conclusions in the non-covariant experiment. The authors documented that the students had great difficulty interpreting non-covariant data. To make up for a lapse in understanding, many students in the non-covariant study attempted to make up “logical reasoning” for a lack of covariation and attempted multiple trials, often replacing specific data points from their original data sets with new findings in an attempt to create a correlation between the two variables. Although students were very confident in their ability to analyze the relationship between two variables that showed a clear relationship to one another, they were often unable to comprehend situations where patterns were not obvious. For students to make conclusions based on scientific reasoning (i.e. reasoning drawn from sound evidence) as opposed to logical reasoning (i.e.: reasoning drawn from non-data evidences such as personal claims), they must be able to analyze data that does not always show obvious patterns.

Finally, in their research exploring assessment practices for learning, Ruiz-Primo and Furtak (2006) found that the incorporation of formative assessments to be essential in the effective science inquiry instruction. Three middle school teachers were selected for their diversity of informal formative assessment practices. These teachers were asked to videotape their classroom sessions, which were then provided to the researchers to be transcribed, coded, and analyzed according to the “ESRU” cycles. In

this cycle, the teacher Elicits a question; the Student responds; the teacher Recognizes the student's response; and then the teacher Uses the information collected to support student learning. The researchers in this study found that the teacher who most closely modeled the ERSU style of informal formative assessment during classroom discussion had the highest levels of student achievement and performance on the embedded assessments. Their results further suggest that the content knowledge of an instructor may have less of an impact on student achievement than assessment strategies that they utilize in the classroom.

Similar to the findings of Lovett and Chang (2007), Keiler (2007) found that students required explicit direct instruction on how to not only conduct data analyses, but furthermore as to the importance or value of these skills. Furthermore, Keiler (2007) also asserted that comprehensive data analysis instruction and assessment incorporates science integration with mathematics. Germann and Aram (1996) went further to describe how these explicit instructions should also be provided to the students in a scaffolded format, and with multiple opportunities for students to interact with data analysis. Hug and McNeill (2008) and Kanari and Millar (2004) both suggest that students need multiple opportunities to interact with data in multiple formats, including first and second hand data sources as well as covariant and non-covariant data types. Finally, Ruiz-Primo and Furtak (2006) suggest that formative assessment strategies should be utilized for effective inquiry-based science education.

Science inquiry is an effective means of providing students with the practice of higher order cognitive skills that they will need to be successful problem solvers as



adults. As many students specifically struggle with the higher order cognitive category of data analysis, science inquiry instruction can be a highly effective tool for aiding students in achieving the skills required for data analysis. There are a variety of instructional strategies that can be implemented to aid science inquiry instruction for data analysis skills to be effective: direct/explicit instruction, science and mathematics integration, scaffolding, multiple data sources, and multiple data types.

### **Summary**

Experiencing science should not be a linear process, but rather a circular, iterative process that emphasizes higher order questioning and thinking. By stepping away from the more commonly practiced “scientific method” and instead focusing on science inquiry, educators are providing students with an engaging and authentic experience with science (Windschitl, Thompson, and Braaten, 2008). Hofstein et al. (2005) assert that students need to have multiple experiences with the science inquiry process in which to practice their higher order thinking skills, opportunities that include both confirmatory and open-ended responses. Through their extensive research synthesis of 18 years of science inquiry education, Miner, Levy, and Century (2009) found the instructional practices that “emphasized student active thinking and drawing conclusions from data” had the strongest positive influence on K-12 student science conceptual learning. Furthermore, Miner, Levy, and Century (2009) found evidence that effectively-implemented science inquiry “intentionally builds on and develops students’ curiosity, enthusiasm, and concentration.” To implement scientific inquiry in the

classroom, the authors recommend three “highly effective” teaching strategies: 1) connecting the inquiry process to science content, 2) utilizing student engagement strategies around science content, and 3) asking students to take responsibility for their learning, thinking, and/or motivation. Zohar and Dori (2003) found that both low and high achieving students made gains in learning when participating in inquiry activities. In fact, Zohar and Dori (2003) found that lower achieving students had higher gains in learning than their higher achieving counterparts when the variable being assessed was specifically their higher order thinking skills. Implemented effectively, science inquiry in the classroom can help students to make gains in higher order questioning and thinking.

Of the many higher order thinking skills that can be improved through the implementation of science inquiry, gaining competence in data analysis is a common struggle for students. Germann and Aram (1996) identified a general unfamiliarity with the conventions of science as an all-encompassing factor in students’ poor overall science performances. Hug and McNeill (2008) characterized four common areas of student difficulty in data analysis, including students’ ability to manipulate data as well as their ability to draw patterns/make inferences from data sets. Lovett and Chang (2007) found that students often struggle with processing observations into organized and useful information that can be applied to future decisions. If students are not provided high performing instruction on how to perform essential tasks, then the instructor is not aiding the students in fully accessing the skill set necessary for their success. Lovett and Chang (2007) thus concluded that for students’ to obtain the necessary skills, instruction in these processes must be made explicit and feedback of student performance needs to

be provided. With these common challenges as well as potentially effective solutions identified, improving students understanding of the essential skill of data analysis is possible.

To effectively and efficiently teach data analysis skills in the classroom, students must interact with a variety of both instructional techniques and data types. Hug and McNeill (2008) identified the need to utilize multiple sources of data in the classroom. Kanari and Millar (2004) found that incorporating both covariant and non-covariant experimentation practices aided in the development of scientific reasoning. Keiler (2007) identified how cooperation between science and mathematics teachers can result in students benefiting from both the planning and evaluation skills used in science class with the analysis and interpretation skills used in math. Finally, Germann and Aram (1996) assert that instructional practices must be scaffolded, and made in conjunction with well-designed rubrics that can assess students higher order cognition. Ruiz-Primo and Furtak (2008) report that formative assessment is an essential ingredient for encouraging higher order thinking and learning in the science inquiry classroom. Through providing students with multiple experiences with a variety of data types as well as high performingly scaffolded and assessed lesson plans, we can provide students with the opportunity to increase their sophistication in data analysis.

### **Research Question**

What are the effects of teacher instructional practices on high school students' ability to analyze science inquiry data? To answer this question, this mixed methods

case study first investigated what instructional practices were implemented as documented by the Teacher Instructional Portfolio. Next, trends were identified in students' ability to analyze data in their work samples. Finally, the study examined any relationships between instructional practices and students' ability to analyze data. The independent variable of this study was the teaching strategies that participating teachers implemented and documented in their units of science inquiry instruction. The dependent variable was the students' ability to analyze data as observed in their student work samples.

## **Methods**

### **Methods Overview**

What effects, if any, do teacher instructional practices have on high school students' ability to analyze data in science inquiry? The goals of this mixed-methods multiple case study were to:

1. Investigate what instructional strategies were evident in the Teacher Instructional Portfolios (TIPs) for supporting high school students in analyzing their data,
2. Examine high school students' ability to analyze data through the Student Work Samples (SWSs) they produced, and
3. Identify possible connections between the teachers' instructional strategies and their student's data analysis abilities.

The teacher instructional portfolios and student work samples of eight high school science inquiry units were analyzed. Quantitative analyses of the TIPs were conducted utilizing the TIP rubric. From full class sets of SWSs, a random selection of eight student work samples was reviewed and scored using the Analyzing and Interpreting Results category of the Oregon Department of Education Science Inquiry Rubric (Oregon Department of Education, 2011). Qualitative data were gathered on both the TIPs and SWSs using researcher developed coding schemas. Analyses of the quantitative and qualitative data were conducted to determine whether there is a

relationship between the teacher instructional practices and the students' ability to analyze data.

## **Participants**

Teacher Participants. Teachers in this study participated in the PSU Connects2Moving Science Education Forward Professional Development and Research Project. As part of a Mathematics and Science Partnerships grant funded project, teachers agreed to 1) participate in a Fall collaboration workshop, 2) create, implement, and document one unit of science inquiry instruction, and 3) participate in a Spring calibration workshop. Each teacher participant in this study was a high school science teacher at an urban or suburban public high school during the time of study. Teacher demographics including sex, gender, race, highest degree earned, teaching experience, years at their current school, and class size can be found in Table 1. The only requirement to participate in this study was enrollment and consent in the PSU Connects2Moving Science Education Forward Professional Development and Research Project.

**Table 1.** *Demographics, experience and education of participating teachers from SEC.*

Class-room*	Gender	Ethnicity	Teaching Experience (years)	Years Teaching at Current School	Highest Degree	Major Field of Study for Bachelor's Degree	Major Field for Highest Degree
Abigail	F	White	6-8	3-5	Master's	Science	Sci Ed
Cathy	F	White	6-8	<1	Master's	Science	Other
Felicia	F	White	1-2	1-2	Multiple Master's	Science	Sci Ed
Joe	M	White	<1	<1	Multiple Master's	Science	Sci Ed
Margot	F	White	6-8	6-8	Multiple Master's	Science	Sci Ed and Sci
Penny	F	White	<1	<1	Multiple Master's	Science	Sci Ed
Sonia	F	White	6-8	6-8	Master's	Science	Sci Ed
Valerie	F	Mixed	>15	12-15	Multiple Master's	Science	Sci

\*The names assigned to the classroom teachers are pseudonyms, here and throughout the study.

Student Participants. The only requirement to participate in this study as a student was enrollment in a participating teacher's science class and parental consent. The 196 SWSs submitted for this study all come from student participants who were enrolled in one of the teacher participants' high school science class sections. As seen Table 2, the students came from diverse racial and ethnic backgrounds and there were a relatively equal number of male and female students among all of the sections. From each class section, eight SWSs were chosen by simple random sampling for analysis.

**Table 2. Student Demographics**

Class- room	Number of students <sup>s</sup>	% ELL <sup>s</sup>	% female <sup>d</sup>	% White <sup>d</sup>	% Black <sup>d</sup>	% Latin o(a) <sup>d</sup>	% Asian / Pacific Island er <sup>d</sup>	% Indi an / AK Nati ve <sup>d</sup>	% Mul ti- eth nic <sup>d</sup>
Abigail	≤ 15	≤ 10	70	38	0	8	31	23	0
Cathy	≥ 26	≤ 10	49	70	3	11	16	0	0
Felicia	16 – 25	11-30	49	54	12	0	18	11	6
Joe	≥ 26	≤ 10	37	66	9	11	11	3	0
Margot	≥ 26	≤ 10	54	50	4	8	38	0	0
Penny	≥ 26	≤ 10	50	59	3	24	12	3	0
Sonia	16 – 25	≤ 10	82	57	0	8	31	23	0
Valerie	No data	11-30	53	23	17	20	33	2	5

(s) - Data obtained from SEC.

(d) - Data provided by the districts.

### Data Collection Instruments.

Survey of Enacted Curriculum (SEC). Each teacher who participated in this study also completed an SEC (Tables 1 and 2). This online survey, developed by University of Wisconsin's Center for Education Research, is research-based data collection and reporting tool designed to provide information about what instructional strategies they are using. The instructional activities questions focus on what student activities (whole group/small group discussion, problem-solving, homework assignments, assessment practices, laboratories, use of educational technology, etc.) are occurring in the classroom and at what frequency. Teacher characteristic questions inquire into what beliefs the teachers holds regarding instructional influences, school climate, and formal educator preparation as well as the effects of years teaching, degrees held, and teaching



certifications. Finally, the SEC asks teachers about the frequencies, amount, and type of professional development they have engaged in during most recent school year.

Teacher Instructional Portfolio (TIP). The TIP is a performance-based instrument that is composed of three main components: 1) TIP Instructions/Prompts, 2) teacher responses to the TIP Tasks and Prompts, and 3) TIP Rubric. The section instructions for Part 1 can be found in Appendix A and the TIP Rubric in Appendix B.

For the TIP Instructions/Prompts section, each participating teacher was provided an empty binder with three sub-sections: 1) Knowledge, Skills, Experiences Outcomes (KSEO), 2) Assessments, and 3) Pedagogical Strategies. In the KSEO section, teachers were asked to provide their learning goals and objectives for the unit, copies of materials utilized in the teaching of the unit (calendar, handouts, PowerPoint slides, text materials, etc.), and other information pertinent to the implementation of the unit. The Assessment section is where teachers documented their formative and summative assessments, as well as examples of student work products. Finally, the pedagogical strategies section is where teachers were prompted to provide reflections.

The teachers' responses to the TIP instructions/prompts were returned to the researchers with each binder sub-section filled with artifacts of their instruction. The TIP rubric was used to quantitatively assess what instructional practices were implemented as documented by the teacher responses to the TIP instructions/prompts. The three domains are 1) classroom roles, 2) content knowledge and cognitive skills, and

3) assessment for learning practices. The multifaceted TIP instrument's validity is currently under investigation by Saxton (pre-publication).

Of the eight teacher participants in this study, only seven submitted TIPs that were deemed scorable against the TIP Rubric. Valerie was ill during the majority of time that would have been used for the implementation of her instructional unit, and the TIP that she submitted was a compilation of instructional materials that she would have liked to use over the year. Thus, Valerie's TIP was deemed unscorable and removed from this study.

Instructional Practices: Qualitative Coding Scheme. In addition to the quantitative assessment of the TIP Rubric, the researchers developed a TIP qualitative coding scheme based on the works of Germann and Aram (1996), Hug and McNeill (2008), Lovett and Chang (2007), Kanari and Millar (2004), Keiler (2007), and Ruiz-Primo and Furtak (2007). This coding scheme is broken into three main coding areas: 1) general instruction, 2) data analysis specific instruction, and 3) data analysis best practices. While the quantitative TIP Scoring Guide provides a broad view of the instructional goals and outcomes of the unit, the TIP qualitative coding scheme allowed the researches to additionally investigate instructional techniques for data analysis skills specifically (Appendix C).

According to the findings of Keiler (2007) and Lovett and Chang (2007), data analysis instruction should be explicit and direct. Germann and Aram (1996) found that

successful data analysis instruction should include forms of embedded scaffolding to help students succeed. Finally, Ruiz-Primo and Furtak (2007) found that successful science inquiry cycles should include formative assessments. For these reasons, TIPs were examined for their incorporation of direct instruction, scaffolding, and formative assessment practices. The TIPs were first evaluated on the general use of direct instruction, scaffolding, and formative assessment practices (“general instruction practices”), and were then evaluated for the specific use of direct instruction, scaffolding, and formative assessment practices for encouraging data analysis (“data analysis specific instruction practices”). Finally, based on the works of Keiler (2007), Hug and McNeill (2008), and Kanari and Millar (2004), a final code was added to the coding scheme which included “data analysis best practices.” These final codes included instructional practices that included STEM collaboration, the use of multiple data sources, and the use of multiple data types.

Student Work Sample (SWS). Each SWS is a performance-based artifact that is composed of three main components: 1) SWS Instructions/Prompts, 2) Student Responses to the SWS Instructions/Prompts, and 3) Oregon Department of Education (ODE) Science Inquiry Scoring Guide: Analyzing and Interpreting Data Rubric (ODE, 2011).

Each participating teacher generated SWS Tasks/Prompts that their students used as the basis to formulate their student responses. The ODE Science Inquiry Scoring Guide: Analyzing and Interpreting Data Rubric was used to quantitatively assess each

student response to the SWS Instructions/Prompts against the science education benchmarks set by ODE. The ODE Science Inquiry Scoring Guide: Analyzing and Interpreting Data Rubric is a single and final section of the four part ODE Science Inquiry Scoring Guide. This ODE Science Inquiry Scoring Guide: Analyzing and Interpreting Data Rubric can be found in Appendix D. The entirety of this multifaceted SWS instrument's reliability is currently under investigation by Saxton (pre-publication).

SWS Data Analysis: Qualitative Coding Scheme. In addition to the quantitative assessment of the ODE Science Inquiry Scoring Guide: Analyzing and Interpreting Data Rubric (ODE, 2011), the researchers developed a qualitative coding scheme adapted from the works of Germann and Aram (1996) and Hug and McNeill (2008) which can be found in Appendix E.

Although the quantitative ODE Science Inquiry Scoring Guide: Analyzing and Interpreting Data Rubric (ODE, 2011) provided a summative score assessing students data analysis abilities, the SWS qualitative coding scheme allowed the researchers to specifically investigate students' data analysis abilities and weaknesses. The coding categories were based on the findings of Germann and Aram (1996) who found the three main areas of student struggle in data analysis to be 1) analyzing results, 2) drawing conclusions, and 3) constructing an argument. The evaluation criteria were based on the work of Hug and McNeill (2008) that identified main areas of student struggle to be their identification of data source, discussion of data agreement,

formation of conclusions, discussions of limitations of data, incorporation of content knowledge, and identification of patterns. These codes provided the researchers with an opportunity to elaborate on findings of the quantitative instrument, further detailing the main areas of student struggle with data analysis.

## **Procedure**

Part 1: Data Collection. During the fall of 2010, participating teachers met to collaborate on pre-existing units of instruction that would later be documented as with the TIP instrument. During the winter of 2011, participating teachers began implementing and documenting their classroom practice and strategies in their TIPs as documented in Table 3. Teachers implementing these science inquiry units also collected and submitted a complete class-set of SWSs. Both the TIPs and the SWSs were submitted to the researchers in the spring of 2011.

Part 2: Coding and Assessment. From spring 2011 to spring 2012, each TIP and SWS was scanned to an electronic filing system. These electronic copies were then redacted of all identifying information. Based on the work of Onwuegbuzie and Leech (2007) using the method of small population statistical analysis, eight SWSs were randomly selected from each set of full class sets of student work samples. These eight SWSs were set aside to be used for final, official data analysis purposed. The remaining student work samples from each class set were used to calibrate scoring among the research group members using the ODE Science Inquiry Scoring Guide: Analyzing and

Interpreting Data Quantitative Rubric. After three rounds of training, the researchers achieved an SPSS Single Measure Intra-class Correlation Coefficient of 0.76 using the ODE Science Inquiry Scoring Guide: Analyzing and Interpreting Data Rubric (ODE, 2011). When satisfactory inter-rater reliability was achieved, the originally retained eight SWS were both quantitatively and qualitatively assessed.

**Table 3.** TIP Science Inquiry Topics, by classroom

Class-room	Grade Level	Content Area	Unit Topic	Abbreviated prompt or task idea
Abigail	11	IB* Biology	Cells	Osmosis Lab - determining the concentration of solutes in a stick of celery
Cathy	10	Chemistry**	Ecology	Plant growth inquiry – The effect of changing conditions on the sprouting and growth of turnip seeds
Felicia	9	Chemistry	Chemical Reactions and Rates	Factors affecting Alka-Seltzer reaction rates: temperature, surface area, and/or concentration
Joe	10	Chemistry	Chemical Reaction and Rates	Factors affecting Alka-Seltzer reaction rates: temperature, surface area, and/or concentration
Margot	11	Physics	2D Motion	Projectile Motion Lab- predicting the motion of a ball rolling of a counter
Penny	9	Biology	Ecology	Plant growth inquiry – The effect of changing conditions on the sprouting and growth of turnip seeds
Sonia	12	IB* Biology	Anatomy and Physiology	Human Heart Rate Lab - Relationship between changes in the cardiovascular system relative to changes in body position or stimuli.

\* IB stands for “International Baccalaureate”

\*\* Teacher reported her content area as chemistry, but unit submitted was life science.

Similarly, four middle school TIPs were selected for the purposes of training and calibrating the raters. These four TIPs were evaluated individually by each rater with the TIP Scoring Guide. After, the scores were discussed as a means of calibration. After the practice evaluation of the four middle school TIPs was complete, the eight high school science inquiry TIPs were evaluated individually by each rater against the TIP Scoring Guide. In the category of Classroom Roles, the researchers were able to achieve an SPSS Single Measure Intra-class Correlation Coefficient of 0.47. In the category of Content and Cognition, the researchers were able to achieve an SPSS Single Measure Intra-class Correlation Coefficient of 0.27. The researchers experienced the greatest level of struggle in the category of Assessment for Learning, where they were only able to achieve an SPSS Single Measure Intra-class Correlation Coefficient of 0.17. The overall Summative SPSS Single Measure Intra-class Correlation Coefficient was 0.63. For this reason, the researchers met to resolve consensus scores in each category of the eight TIPs included in this study. Following the quantitative analysis, further qualitative analyses were also conducted on each TIP.

Part 3: Data Analysis. In the winter of 2012, quantitative and qualitative data were analyzed. Case descriptions for each classroom were written to frame the context of instruction and learning. Both the TIP and SWS quantitative rubrics generated individual numerical scores, as well as a composite score for each classroom. The TIP and SWS qualitative scores generated frequency data for each classroom. TIP and SWS

quantitative scores were compared against one another to investigate trends. Qualitative data were used to further illustrate what instructional strategies were used to support data analysis skills, the data analysis skills exhibited by the students, and to determine whether any of the instructional strategies were associated with student outcomes with regard to data analysis abilities.



## Results

This study was designed to investigate the effects of teacher instructional practices on high school students' ability to analyze science inquiry data. To examine this, we first investigated what instructional practices were implemented as documented by the Teacher Instructional Portfolio. Next, trends in students' ability to analyze data in their work samples were identified. Finally, relationships between instructional practices and students' ability to analyze data were examined. To clarify the context of each unit of instruction, a case descriptions of each class was developed.

### Case Descriptions

Abigail. Abigail is a white female with 3-5 years of experience. She holds a master's degree in science education and a bachelor's degree in science. Her certifications are in both secondary and middle level science, and she has been teaching at her current school for 3-5 years. She currently teaches 11<sup>th</sup> grade IB Biology to a class of less than 15 students, the smallest class in this study. Her students are predominantly female (69.2%), and come from a variety of ethnic backgrounds including White (38.5%), Asian/Pacific Islander (30.8%), Hispanic (7.7%), and Native American/Alaskan Native (23.1%), and 10% or less of her students are English Language Learners.

The five-week unit she prepared for this research study covered topics in cell biology, including cell structure and function, as well as cell processes, including passive

transport and mitosis. The student science inquiries investigated osmosis by determining the solute concentration of a celery stick. The inquiries were relatively open-ended, no template was provided, and the students were active participants in the classroom. The labs were introduced before the lectures, thus providing the students with an opportunity to experience the content via discovery. The teacher routinely assessed students for learning and understanding, but the researchers could not ascertain whether Abigail used this information to inform her instruction. On the handouts and other seatwork, the questions tended to be lower order and the overall emphasis was on content knowledge. Higher-order questions were present, but implementation was inconsistent over the course instruction.

Cathy. Cathy is a white female with 6-8 years of experience. She holds a master's degree in a non-science education discipline but her bachelor's degree is in science. Her certifications are in both secondary and middle level science, and she has been teaching at her current school for 3-5 years. She currently teaches a 10<sup>th</sup> grade class of more than 26 students. Although Cathy listed her course title as Chemistry on her SEC, the unit she submitted appeared to contain content more closely aligned with a Life Science course. Her students were mostly White (70.3%) and male (51.4%), and 10% or less of her students are English Language Learners. Her non-White student population was composed of ethnic groups including African American (2.7%), Asian/Pacific Islander (16.2%), and Hispanic (10.8%).

The unit she prepared for this research study was on ecology, and the student science inquiries investigated turnip seed growth. The topics covered over the course of this 17 day unit were energy transfer through ecosystems, ecosystem relationships, matter cycles, and the impact of humans on ecosystems. The instructional practices in this unit emphasized collaborative efforts and group work. The labs emphasized the use of science inquiry templates as scaffolding. Her warm-up questions encouraged concept application and higher order thinking, but many of the handouts focused on lower order rote memorization style drills. According to her pedagogical reflection, Cathy's students spent 40% of class time participating in group work, 40% participating in their inquiry experiments, 15% listening to lectures, and 5% in class discussion. Cathy was one of only two teachers in this study (the other was Penny) who documented three full class periods dedicated to data analysis. Cathy also routinely assessed students for learning and understanding, but again we could not ascertain whether this information influenced her instruction.

Felicia. Felicia is a white female with 1-2 years of experience. She holds multiple master's degrees in science education and a bachelor's degree in science. Her certifications are in both secondary and middle level science as well as an emergency or provisional license, and she has been teaching at her current school for less than 1 year. She currently teaches 9<sup>th</sup> grade physical science to a class of between 16 and 25 students, in which 11-30% have limited English language proficiency. Her student body

was 51.4% male and 54.0% White. Her non-White student population was composed of ethnic groups including African American (12.2%), Asian/Pacific Islander (17.6%), Multiracial (5.4%), and Native American/Alaskan Native (10.8%).

The unit she prepared for this research study was on chemical reactions, in which the student science inquiries investigated reaction rates of Alka-Seltzer in water with varied temperatures or surface areas, with a few students electing to investigate both simultaneously. The unit was five weeks long, with three weeks spent on balancing reactions, one week spent on reaction rates, and the final week dedicated to the scientific inquiry. According to Felicia, 60% of class time was spent in group work, 20% was spent in class work or practice, 5% in direct instruction, and 5% in homework. It should be noted that this only adds up to a total of 90%. The students were provided with scaffolding in the form of optional templates, and overall this unit emphasized active student inquiry and participation. In fact, this was one of few units where content appeared to be less emphasized than cognitive skills. This teacher also routinely provided her students with assessment opportunities, but again there was no evidence regarding whether Felicia used the assessment data to influence her instruction. Peer assessment was emphasized.

Joe. Joe is a white male with less than 1 year of experience. He holds multiple master's degrees in science education and a bachelor's degree in science. His certification is in secondary level science. He currently teaches 10<sup>th</sup> grade Chemistry to

a class of more than 26 students with 10% or less of students having limited English language proficiency. His students were predominantly male (62.9%) and White (65.7%). His non-White student population was composed of ethnic groups including African American (8.6%), Asian/Pacific Islander (11.4%), Hispanic (11.4%), and Other (2.9%).

The unit he prepared for this research study covered chemical reactions, and the science inquiries investigated reaction rates of Alka-Seltzer in water with varied temperatures and/or surface areas. The unit was five weeks long and covered stoichiometry, limiting reactants, and factors that influence reaction rates. According to Joe, 30% of class time was spent in discussion, 20% in group work, 20% in demonstrations, 10% in lecture, 10% in group data collection, 5% in peer-peer reviews, and 5% watching videos. This unit was science inquiry driven, with many labs and student discussion. Although a template was not provided to the students for their inquiries, direct instruction and notes were provided to students emphasizing Joe's desired outcomes. Students were provided with multiple opportunities to observe and explain, providing a balanced emphasis between content and cognitive skills. Joe utilized diverse forms of assessment tied to specific learning targets, as well as opportunities for students to provide each other with peer feedback.

Margot. Margot is a white female with 6-8 years of experience. She holds multiple master's degrees in science education and a bachelor's degree in science. Her

certifications were in both secondary and middle level science, and she has been teaching at her current school for 3-5 years. She currently teaches 11<sup>th</sup> grade physics to a class of more than 26 students with 10% or less of students having limited English language proficiency. Her student body consisted of 53.8% females, and come from a variety of ethnic backgrounds including White (50.0%), African American (3.8%), Asian/Pacific Islander (38.5%), and Hispanic (7.7%).

The unit she prepared for this research study covered two-dimensional motion, and the student science inquiries specifically investigated projectile motion. This unit lasted 23 days and emphasized students deriving equations for two-dimensional kinematics that include gravity and projectile motion. The calendar breakdown was as follows: 14 days were spent in lecture, watching videos, or viewing online simulations; 2 days were spent graph matching with Vernier motion sensors; 1 day was spent on independent student practice with the Vernier probes; 4 days on practice tests and testing; and the last two days were dedicated to the science inquiry. The science inquiry asked students to compare theoretical and experimental results for the distance a projectile lands when rolling off a countertop from a given height.

Margot openly stated that she emphasized direct instruction practices throughout the course of this unit. According to her pedagogical reflection, 50% of time was spent in lecture, 25% was spent with students working independently, 12% group work, 8% lab time, and 4% using simulators. The time spent in the inquiry laboratory was relatively student directed. No templates were provided to the students, but

scaffolding in the form of expectation handouts was used. Although Margot provided her students with two opportunities to conduct science inquiry style labs, the reports seemed to emphasize the student's ability to manipulate equations and not necessarily understand them. Furthermore, Margot used many homework assignments to encourage student learning, but it did not appear that the homework assignments were checked for anything more than completion (in the form of a stamp). Thus, while the content in this unit appeared developmentally appropriate, the lack of emphasis on cognitive skills did not.

Penny. Penny is a white female with less than 1 year of experience. She holds multiple master's degrees in science education and a bachelor's degree in science. Her certification is in secondary level science. She teaches 9<sup>th</sup> grade Life Science to a class of more than 26 students with 10% or less of students having limited English language proficiency. Her student population is half male (50.0%), and 58.8 % White. Her non-White student population was composed of ethnic groups including African American (2.9%), Asian/Pacific Islander (11.8%), Hispanic (23.5%), and Native American/Alaskan Native (2.9%).

Like Cathy, the unit Penny prepared for this research study was on ecology, and the student science inquiries investigated turnips seed growth. Again, the topics covered over the course of this 17 day unit were energy transfer through ecosystems, ecosystem relationships, matter cycles, and the impact of humans on ecosystems.

Unlike Cathy, Penny did not emphasize collaborative learning and instead spent approximately 40% of the unit utilizing direct instruction, 30% in independent inquiry, 10% on class discussion, 10% on cooperative learning, and 10% in pair work. Many power point slides were used, but an overall lack of cohesiveness was predominant throughout this unit. While the overall unit was “ecology” the experiment the students conducted investigated plant growth and the connection between these two concept strands was never made apparent. Both Cathy and Penny used excellent warm-up questions to encourage higher order thinking, but Penny provided students with more clear opportunities to connect content with cognitive skills. In her pedagogical reflection, Penny said that she routinely assessed her students, but outside of this mention of conversational feedback there was no evidence regarding her use of this information to influence instruction. Finally, Penny was the only other teacher to dedicate three full class periods to data analysis.

Sonia. Sonia is a white female with 6-8 years of experience. She holds a master’s degree in science education and a bachelor’s degree in science. Her certification is in secondary level science, and she has been teaching at her current school for 3-5 years. She currently teaches 12<sup>th</sup> grade IB biology to a class of about 16 to 25 students with 10% or less of students having limited English language proficiency. Her students were predominantly female (82.6%) and White (56.5%). Her non-White student population



was composed of ethnic groups including Asian/Pacific Islander (34.8%) and Hispanic (8.7%).

The unit she prepared for this research study was a 28 day anatomy and physiology unit. The specific content covered was the digestive, respiratory, and circulatory systems. The student science inquiries investigated the relationship between changes in the cardiovascular system relative to changes in body position or stimuli. According to the teacher, 26% of time was spent in pair work, 25% in conducting individual inquiries, 23% on videos and class discussion, 22% in direct instruction, and 4% in cooperative learning. While student participation in this unit was primarily passive, the labs built progressively on one another. The first labs were scripted with provided procedures, but these directions were slowly released, encouraging the development of a synthesized set of cognitive skills. This unit built from simple, scaffolded labs to a summary science inquiry that was both complex and open ended. Like the units above, there was no evidence that the results of this unit's assessments were used to inform instruction, but the concepts covered were developmentally appropriate and peer assessments were used.

### **Teacher Instructional Portfolio Results**

Quantitative TIP Results. As seen in Table 4, only two of the seven (28.6%) TIPs scored an high score of 3 in all three categories: Classroom Roles, Content and Cognitive Skills, and Assessment for Learning. Two of the seven (28.6%) TIPs scored low in one of

the categories. Another two of the seven (28.6%) TIPs scored low in two of the categories, and one of the seven (14.3%) of TIPs scored low in all three of the categories.

Four of the seven (57.1%) TIPs received high score of 3 in Classroom Roles, while the remaining three (42.9%) received a low score of 2. Similarly, four of the seven (57.1%) TIPs received a high score of 3 in Content and Cognitive Skills, but the remaining three (42.9%) received low score of 2. Finally, four of the seven (57.1%) TIPs received high score of 3 in Assessment for Learning, but the three (42.9%) received low score of 2.

**Table 4.** *Summary of TIP Quantitative Scores*

Class ID	Abigail	Cathy	Felicia	Joe	Margot	Penny	Sonia
Classroom Roles	3	3	3	3	2	2	2
Content and Cognitive Skills	2	2	3	3	2	3	3
Assessment for Learning	3	2	3	3	2	2	3
Summative Total	8	7	9	9	6	7	8

Qualitative TIP Results. In general, the TIPs scored relatively high with regard to their implementation of data analysis instruction (Table 5). Only one of the seven TIPs included in this study failed to include data analysis instruction in its documentation of the unit design.

Six of the seven TIPs included in this study documented the use of direct instruction practices for data analysis skills. Four TIPs received a score of 2, meaning that they explicitly targeted data analysis. Two TIPs received a score of 1 for direct

instruction that vaguely mentioned data analysis, meaning that the reference to data analysis was not explicit. All TIPs included in this study received a score of 1 for vague reference to formative assessment of general instruction and a score of zero for their formative assessment of data analysis concepts.

**Table 5. Classroom Results of TIP Qualitative Scores**

	Abigail	Cathy	Felicia	Joe	Margot	Penny	Sonia
<b>General Instructional Practices</b>							
Direct Instruction	2	2	2	2	2	2	2
Formative Assessment	1	1	1	1	1	1	1
Scaffolding	0	2	2	2	2	2	2
<b>Data Analysis Specific Instructional Practices</b>							
Direct Instruction of Data Analysis	0	2	2	2	2	2	2
Formative Assessment of Data Analysis	0	0	0	0	0	0	0
Scaffolding of Data Analysis	0	2	1	2	1	2	2
<b>Data Analysis Best Practices</b>							
STEM Integration	0	0	0	0	1	0	0
Multiple Data Sources	0	0	0	0	0	0	0
Multiple Data Types	0	1	0	0	1	1	1
Summative Instruction and Data Analysis Total	3	10	8	9	11	10	10

Six of the seven TIPs documented the use of scaffolding techniques in their instructional practice. Abigail's TIP, the only one that did not use scaffolding, was also the only TIP that did not document any data analysis specific instructional practices. Six of the seven TIPs also documented the use of scaffolding for data analysis concepts in their instructional practice.

As seen in Table 5, the TIP from Abigail received the lowest qualitative summative total for both general instructional practices and data-analysis-specific instructional practices. Felicia and Margot's TIPs received a qualitative summative total of 3 for their data-analysis-specific instruction, while the TIPs from Cathy, Joe, Penny, and Sonia received the highest qualitative summative total of 4 for their data-analysis-specific instruction.

While no teachers in this study explicitly incorporated data analysis best practices, four teachers did vaguely incorporate two of the data analysis best practice instructional techniques. Margot included practices aligned with the integration of math and science during her Physics inquiry on projectile motion. Cathy and Penny both vaguely included multiple data types during their turnip seed investigations, as did Sonia in her cardiovascular system inquiry. No teachers in this study utilized multiple data sources.

Margot received the highest score for her data analysis instruction, a score of 6. Overall, Abigail's, Felicia's and Joe's TIPs received the lowest qualitative summative totals for both instruction and data analysis, with scores ranging from 3-9 points. Cathy, Margot, Penny, and Sonia's TIPs received a high qualitative summative totals ranging from 10 to 11 points.

## Student Work Sample Results

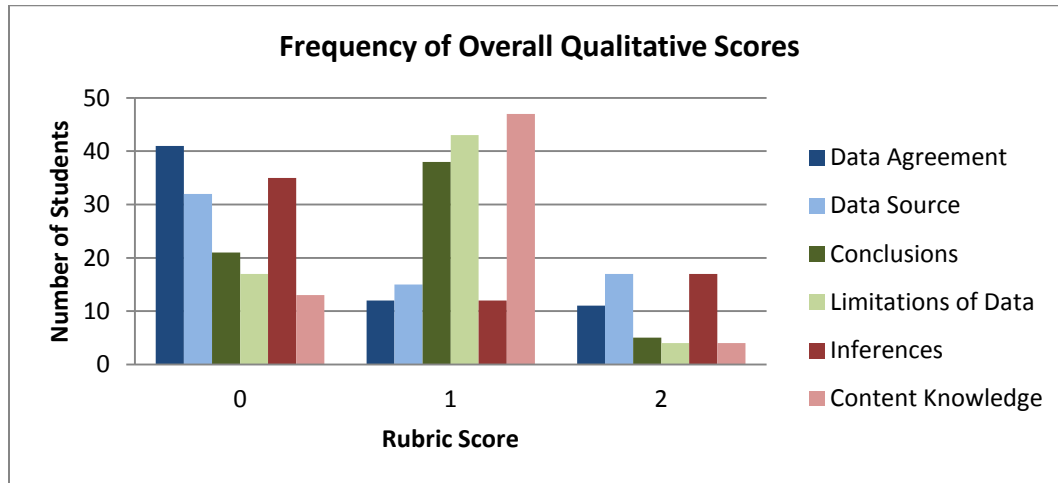
Quantitative SWS Results. Table 6 shows the quantitative SWS results by class. The highest scoring classes in this study belonged to Abigail and Joe, whose students averaged a score of 4 on their SWSs. The five remaining classrooms (belonging to Cathy, Felicia, Margot, Penny, and Sonia) received average scores of 3, which identifies their analyses as not proficient.

**Table 6.** *Quantitative SWS Results by Classroom*

	Abigail	Cathy	Felicia	Joe	Margot	Penny	Sonia
Total 1	0	0	2	0	1	0	0
Total 2	1	2	0	0	3	2	0
Total 3	4	5	4	4	3	5	6
Total 4	1	1	2	4	1	1	2
Total 5	2	0	0	0	0	0	0
Total 6	0	0	0	0	0	0	0
Average Score	4	3	3	4	3	3	3

An examination of the overall frequency of student scores across all eight classes against the ODE rubric reveals that the majority (49 of the 64 SWSs, or 76.6%) of students in this study did not receive a passing/proficient score of 4 or higher on their data analyses. Only 15 of the 64 SWSs (23.4%) received proficient scores of 4 or higher. It is also important to note that while three SWS received the lowest score of 1, only 2 SWSs received a score of 5 and no SWSs received a high score of 6.

Qualitative SWS Results. A qualitative analysis of the SWS showed that the majority of students struggled with data analysis skills in the highlighted areas (Figure 1).



**Figure 1.** *Summary of SWS Qualitative Scores*

In the Analyzing Results domain, the greatest areas of student struggle were in the sub-categories of Data Agreement and Data Source. The majority of SWSs (41 of the 64, or 64%) scored a zero for no evidence in the category of Data Agreement. Similarly, half of the SWSs (32 of the 64, or 50%) received a score of zero for no evidence in the category of Data Source.

In Drawing Conclusions, the majority of SWSs (38 of the 64, or 59%) received a score of 1 for vague or unclear evidence with regard to their ability to draw conclusions. Sixty-seven percent of SWSs (43 of the 64) received a score of 1 for vague or unclear evidence for recognition of limitations of the data.

Finally, in Communicating Results, the majority of students struggled with the development of Patterns and/or Inferences (35 of 64, or 54% receiving a score of zero

for no evidence). Similarly, 73% of SWSs (47 of 64) scored a 1 for vague or unclear evidence with regard to their inclusion of Content Knowledge.

As seen in Table 7, SWSs from Margot's class received the lowest qualitative summative total of 19. Joe, Felicia, and Abigail classes received medium scores ranging from 28-35 total points. Classes Cathy, Penny, and Sonia all received the highest marks, ranging from 43-57 total points.

**Table 7. Summative Classroom Results of SWS Qualitative Scores**

	Abigail	Cathy	Felicia	Joe	Margot	Penny	Sonia
Analyzing Results							
Data Agreement	2	7	0	0	0	14	11
Data Source	0	15	6	0	2	11	8
Drawing Conclusions							
Conclusions	6	3	15	10	3	6	5
Limitations of Data	10	6	2	9	5	6	9
Communicating Results							
Patterns/Inferences	8	6	1	1	2	6	15
Content Knowledge	9	6	6	8	7	4	9
Summative Total	35	43	30	28	19	47	57

## Comparative Results

Comparative Quantitative Results. Tables 8-11 compare TIP Scores with SWS Mean Scores by classroom. Tables 8, 9, and 11 all showed a similar relationship that classrooms with low TIP scores more often received low SWS scores, but the same relationship was not evident for high TIP scores and high SWS scores. Table 10, comparing TIP Content and Cognitive Skills Score with SWS Mean was the only comparison revealing a low scoring TIP being associated with high performing scoring

SWS. As can be seen in Tables 8-11, Felicia, Joe, and Margot consistently remain in the same quadrants of the contingency tables.

		SWS Mean	
		Not High performing (1,2,3)	High performing (4,5,6)
<b>TIP</b> Summative Total	<b>Low</b> ( 1,2)	Cathy, Margot, Penny	
	<b>High</b> (3,4)	Felicia, Sonia	Abigail, Joe

		SWS Mean	
		Not High performing (1,2,3)	High performing (4,5,6)
<b>TIP Category 1</b> Classroom Roles	<b>Low</b> ( 1,2)	Margot, Penny, Sonia	
	<b>High</b> (3,4)	Cathy, Felicia	Abigail, Joe

		SWS Mean	
		Not High performing (1,2,3)	High performing (4,5,6)
<b>TIP Category 2</b> Content and Cognitive Skills	<b>Low</b> ( 1,2)	Cathy, Margot	Abigail
	<b>High</b> (3,4)	Felicia, Penny, Sonia	Joe

		SWS Mean	
		Not High performing (1,2,3)	High performing (4,5,6)
<b>TIP Category 3</b> Assessment for Learning	<b>Low</b> ( 1,2)	Cathy, Margot, Penny	
	<b>High</b> (3,4)	Felicia, Sonia	Abigail, Joe



Comparative Qualitative Results. Table 12 compares summative TIP Scores with SWS Scores by classroom. Again no overarching patterns emerged from comparison of the classroom data in this study. A weak relationship between high TIP scores and high SWS scores is observed, but the same relationship was not evident for low TIP scores and SWS scores.

Table 12. Classroom Results by TIP Summative Total versus SWS Summative Total		SWS SUM	
		Low (0-30)	High (31-60)
TIP SUM Instruction and Data Analysis	Low (0-9)	Joe	Abigail
	High (10-18)	Felicia, Margot	Cathy, Penny, Sonia

## Discussion

This multiple case study was designed to investigate the effects of teacher instructional practices on high school students' ability to analyze science inquiry data. First, trends in local high school instructional practices were analyzed with TIPs. Next, trends in students' ability to analyze data identified via their SWSs. Finally, relationships between instructional practices and students' data analysis ability were examined.

### **Trends in Teacher Instruction**

The first question asked by this research study investigated instructional practices teachers are implementing in their classrooms, as documented by the Teacher Instructional Portfolio. The TIP rubric used to score each TIP is composed of three sections, each investigating an essential element of classroom instruction and aligned with the quantitative TIP Scoring Guide. The first section, Classroom Roles, is designed to analyze the nature of instruction with regard to the balance of student and teacher centeredness. The second, Content and Cognitive Skills, explores where instructional emphasis is placed with regard to these two essential education components. The third, Assessment for Learning, investigates what types of and how frequently assessments are being utilized in the classroom. The final two categories discussed in this section are not TIP Rubric sections, but are instead qualitative examinations of what practices were being utilized to encourage learning and data analysis skills as documented by the TIPs.

Classroom Roles. Four of the seven TIPs received high score of three in Classroom Roles, while the remaining three received a high score of two (Table 1). Abigail, Cathy, Felicia, and Joe received the highest scores because the majority of their lessons and activities emphasized student-centered learning, in which students were actively participating and drawing conclusions based on their own observations. These four teachers demonstrated the ability to utilize facilitation techniques that encourage students to take an active role in the learning process. For example, according to Joe's pedagogical reflection, the students in Joe's classroom spent the majority of their time in either class discussions or group work. Some of the activities he used were in the direct instruction style, but many were model-based and asked students to observe demonstrations and then use reasoning to explain their new understandings. Although direct instruction can be an effective strategy in helping students gain conceptual understanding, best practices recommend that a variety of instructional techniques are being utilized to foster student growth and higher order thinking (NRC, 2012). The instructional practices demonstrated in the TIPs emphasize science inquiry skills, where the central authority of knowledge is being transferred away from the teacher and ownership of understanding is presented to the students (Anderson, 2002; Miner, Levy, and Century, 2010).

Three teachers in this study received a low score in Classroom Roles. Margot, Penny, and Sonia's TIPs documented instructional practices that were teacher-centered,

with students as passive participants, and utilized overly scaffolded approaches. In Penny's ecology unit, for example lecture style PowerPoint presentations were used during half of the instructional time. Furthermore, Penny used many videos, one day dedicating an entire class period to watching a video that did not appear to be discussed during the subsequent lessons. Although these teachers did occasionally utilize group work and problem solving activities, the emphasis lay on students relying on their teacher or another source of authority (textbook, video, etc.) to determine the accuracy of their understanding. Presenting material to students, particularly with emphasis on content, does not allow the student to take an active role in their learning (Anderson, 2002; Miner, Levy, and Century, 2010).

With regard to scaffolding, these teachers provided students with mandated templates for their science inquiry projects. Scaffolding can be a very effective strategy for encouraging student learning (Germann and Aram, 1996), but this research project suggests that requiring students to use a template inhibited their responses in their final work samples. Such scaffolding via templates prevented students from developing analyses that were cohesive, as they were broken into isolated categories, thus making it impossible for them to meet the criteria necessary for a 5 or 6 on the ODE Science Inquiry Scoring Guide: Analyzing and Interpreting Data Rubric (ODE, 2011). This was seen in the science inquiry prompts of Sonia's TIP. Sonia's "Discussion of Results" sections were highly scripted, asking students specific, lower-level questions, many of which could be answered with one word responses. Her students were unable to

produce a final product that would score a 5 or higher on the ODE Science Inquiry Scoring Guide: Analyzing and Interpreting Data Rubric, specifically because of the template they followed (ODE, 2011).

Content and Cognitive Skills. Four of the TIPs received high score of three in Content and Cognitive Skills, and the remaining three received low score of two. Teachers receiving a high score in this section (Felicia, Joe, Penny, and Sonia) did so because the curriculum they designed targeted both content and cognitive skills, but more specifically because they provided multiple opportunities for students to construct meaning, apply content, and practice higher-order cognitive skills. An example of this was evident in Sonia's TIP, documenting instruction in her IB Biology classroom. Although Sonia's TIP was considered a level two for Classroom Roles, her scaffolding of inquiry in this lesson earned her a three in Content and Cognitive Skills. This unit began with simple, highly scripted inquiries, but as the inquiry investigations progressed, the scaffolds were removed and the templates provided became increasingly less scripted. By the end of this unit, students were completing open-ended inquiries that required them to draw on not only their observations, but also the content discussed in class to formulate evidence-based conclusions. This implies, in agreement with the results of Hofstein et al. (2005) that students who participate in multiple science inquiry opportunities are more prepared to ask higher order questions. This, and other, research suggest that science classroom experiences should include more opportunities

to experience science inquiry and thus encourage students to participate in higher order thinking (Anderson, 2002; Miner, Levy, and Century, 2010; Windschitl, Thompson, and Braaten, 2008).

Conversely, IB Biology instructor Abigail received low score of two for Content and Cognitive Skills in her classroom. Unlike Sonia, Abigail provided students with no templates for scaffolding and introduced science inquiry labs before lecturing on the topics. The analysis questions of these labs prompted primarily lower order thinking, for example asking questions that elicit one word, memorized response. Although higher order questions were present, and the work produced by Abigail's students was of high quality, the opportunities students had to practice higher order thinking and reasoning skills were incomplete, inconsistent, or infrequent. In agreement with the findings of Anderson (2002), it is clear that students should be provided with multiple opportunities to experience science inquiry in order to surpass a level of understanding beyond that of rote memorization. The scaffolding of these inquiries helps students experience deeper conceptual gains in understanding. As such, scaffolds should be designed as a means of aiding students in their understanding, but not limit their ability to think beyond the scaffold (Ruiz-Primo, Li, Tsai, and Schneider, 2010).

Assessment for Learning. As seen in Table 4, Abigail's, Felicia's, Joe's, and Sonia's TIPs received high score of three in Assessment for Learning. Cathy, Margot, and Penny received low score of two. This category of the TIP Scoring Guide specifically

investigated what type of assessments were being utilized, with what frequency they were used, whether these assessments included self-assessments, and if the cognitive demand of the assessments was developmentally appropriate. Felicia provided a good example of assessment practices. She assessed students' understanding daily via higher order thinking "warm-up" questions, and emphasized the use of peer assessment to encourage self-reflection over the course of her chemical reactions unit. The cognitive demand of the unit assessments was challenging, but completely developmentally appropriate with standards aligned with those recommended by the ODE (ODE, 2012). One area of weakness for Felicia, which is also true for all other teachers in this study, was that she did not provide evidence of if or how she used formative assessments to educate her instructional practice. Teachers in this study were encouraged to document changes they made to the lesson plans and calendars, specifically with regard to their questioning strategies and their observation notes. Of all of the artifacts collected, this appeared to be the one that was most often deficient. As assessment for learning is an essential component for encouraging student growth and development as learners (Black and Williams, 1998), and a critical component for increasing instructional effectiveness (Popham, 1999), further investigation into the insufficient formative assessment findings of this study is necessary.

Three TIPs received low score of two, those belonging to Cathy, Margot, and Penny. Margot's TIP, which scored low in every TIP Rubric section, provided evidence of instruction that did not incorporate high performing assessment techniques. Margot

assigned large amounts of physics homework, but there was no evidence that these assignments received more than a stamp for completion. Similarly, the two inquiry labs conducted in this unit were content driven, asking students to manipulate equations but not necessarily reason or make sense of their applicability. Although the content covered in this lesson was developmentally appropriate, the cognitive skills were not. Her science inquiry was a confirmatory experiment, asking students to calculate predictions and then check to see if they were right via the experiment. Although this could have been a challenging inquiry, the resultant student work did not show that students understood or made sense of the data they collected. This style of assessment moves students away from higher order thinking and can push for rote memorization and the seeking of “the right answer” (Ruiz-Primo and Furtak, 2007; Webb and Jones, 2009).

Finally, this research found evidence that not all instructors provided their students with the ODE Science Inquiry Scoring Guide (ODE, 2011). The students who were not made aware of the scoring guide were at a disadvantage to those students who were. Having knowledge of the rubric, or having an idea of the expectations of the assessment, can aid a student in being able to provide a quality SWS (Andrade, Du, and Wang, 2008). Thus, in future studies the researchers would suggest making it clear to teacher participants that the scoring guide is made available to all students in advance of their final submission.



General Instructional Practices. Qualitative analysis of the TIPs provided further insight into the variety of instructional practices utilized in each classroom. Every teacher involved in this study documented his or her use of direct instruction, most often in the form of power point or lecture style presentations. Although direct instruction is indeed a means of helping students to gain a mastery of content knowledge (Klahr and Nigram, 2004), science inquiry lends a stronger tool to help students gain higher-order thinking skills (Anderson, 2002; Dean and Kuhn, 2006; Germann and Aram, 1996; Hofstein, Navon, Kipnis, and Mamlok-Naaman, 2005; Miner, Levy, and Century, 2009; Windschitl, Thompson, and Braaten, 2008; Zohar and Dori, 2003). It is thus informative that the most commonly held instructional practice observed throughout this study was the use of direct instruction techniques.

Similarly, every teacher documented the use of some type of formative assessment (daily warm ups, exit slips, closure questions, etc.). While a broad spectrum of formative assessment practices exist, this research specifically looked for practices which incorporated feedback or influenced daily instruction (Black and Williams, 1998; Popham, 1999; Sadler, 1989). No teachers in this study provided evidence of how their quick, daily formative assessments influenced their instruction. As previously mentioned, formative assessment is an essential component to instruction, and these omissions should be investigated through further research.

Finally, all but one teacher provided evidence of using scaffolding in their lesson plans, usually in the form of a template provided to students over the course of their

science inquiry labs. The only teacher who provided no evidence of scaffolded instruction was Abigail, who taught one of the two IB Biology sections included in this study. Research suggests that providing students with the option of scaffolding produce greater levels of achievement than providing no scaffolding or mandatory scaffolding (Simons and Klein, 2007). Furthermore, Oregon will soon be adopting the Next Generation Science Standards best practices, which do suggest that at least some scaffolding be provided to students to help them effectively achieve gains in learning (NRC, 2012).

Data Analysis Specific Instructional Practices. Qualitative analysis of the TIP gave the researchers a look into the instructional practices being used to specifically foster data analysis skills. Six of the seven teachers in this study provided evidence that they used some form of direct instruction for data analysis. Abigail was the only teacher in this study who did not provide direct instruction for data analysis skills (Table 5). Although Abigail did not mention data analysis in her TIP, the SWSs her students produced scored very high against the ODE Science Inquiry Scoring Guide (ODE, 2011). This implies to the researchers that Abigail's students were provided with undocumented instruction, either in Abigail's IB Biology classroom or in previous experience conducting science inquiries.

The instructional practices that did incorporate direct instruction for data analysis ranged from a single slide mentioned during a lecture in Felicia's classroom to three days of data analysis discussion in Margot's classroom. The diverse range of direct

instruction implementation for data analysis skills gave researchers insight into the variety of practices being used to improve student learning and comprehension of data analysis skills. Lovett and Chang (2007) found that students required direct instruction not only on how to complete data analyses, but that students benefited from explicit instruction as to the significance of said analysis. While this study produced evidence that direct instruction for data analysis skills was being used in the classrooms studied, no evidence was found that this instruction targeted the significance of data analysis in general.

With regard to formative assessment of data analysis, no teachers involved in this study provided evidence that they formatively assessed their student's data analysis abilities. Although peer-assessments of the SWS were used in Cathy's, Joe's, and Sonia's classrooms, no evidence was provided that these sessions were dedicated to the building of students' data analysis skills and understandings. Thus, a limitation of this study was that the researchers were unable to observe daily classroom conversation or dialogue. While Ruiz-Primo and Furtak (2007) videotaped the ESRU cycles occurring during each individual classroom the studied, this study utilized instruments designed to obtain larger scale information of overall classroom practices. As such, data regarding informal formative assessments may not have been included in this study.

Finally, with regard to scaffolding of data analysis skills, Abigail again provided no evidence of scaffolding. Felicia and Margot each provided evidence of some form of scaffolding, usually in the form of an optional template to be utilized during the science

inquiry investigations. Cathy, Joe, Penny, and Sonia provided evidence of highly scaffolded data analysis instruction, specifically with regard to providing a mandated template used for the science inquiries. According to Ruiz-Primo et al. (2010), a balance must be struck in the amount of scaffolding given to students. Teachers who do not provide enough scaffolding are not aiding their students of the highest need, but providing too much scaffolding can prevent students from being able to attain or express their higher levels understanding. In this study, many students were provided templates as a means of scaffolding.

Data Analysis Best Practices. Although no teachers in this study explicitly incorporated the Data Analysis Best Practices techniques, four TIPs in this study did have vague references to two of the best practice strategies. Cathy, Penny, and Sonia each implemented inquiry experiments that included non-covariant and covariant data, as suggested by Kanari and Millar (2004). Like the results found in the research of Kanari and Millar (2004), many students in these classes struggled to “make sense” of covariant data. The majority of students in these classes attempted to find significance in their results, often by disregarding P-values, as will be further discussed in the qualitative analysis of the SWSs.

Margot’s physics science inquiry was mathematics heavy, having students use advanced algebra to predict projectile motion. Research has found that this practice of integrating math and science has the ability to not only enrich students understanding

of math by giving it an application to a real world experience; it also helps the students by providing them with mathematical tools to analyze and interpret science (Keiler, 2007). Contrary to the previously mentioned research by Keiler (2007), the SWS that resulted from Margot's class were considerably low. It is important to consider that while Margot did incorporate math and science integration, she received the lowest overall TIP score in this study, being considered low performing in all categories of the TIP. This suggests that while STEM integration is important to improving student comprehension across multiple disciplines, this process must be taught in conjunction with other, more general best practices.

Trends in Teacher Instruction Summary. Only two of the teachers in this study used best instructional practices for supporting students in science inquiry across all areas measured by the TIP rubric. These two teachers, Felicia and Joe, specifically provided the researchers with insight into what a unit employing these practices can look like. These two teachers encouraged student-centered learning, provided a curriculum with balanced emphasis on content knowledge and cognitive skills, and they assessed their students for learning. While Joe and Felicia provided strong examples of general instructional practices, neither of these teachers provided an exemplary example of data analysis instruction. Neither teacher provided evidence of using any of the instructional practices encouraged by the literature for the promotion of students' data analysis abilities.

The TIPs submitted by Abigail and Sonia scored similarly, as they were both high performing in two categories of the TIP but each struggled in one section of the TIP. Qualitative analysis of these TIPs found that Sonia utilized a variety of instructional practices in her classroom that encourage data analysis skills, including the data analysis best practice of incorporation of multiple data types. Abigail, on the other hand, was the most deficient teacher in this study with regard to inclusion of instructional techniques for encouraging data analysis skills.

Cathy and Penny each received high score in only one section of the TIP, and were both found to be low performing in the two remaining sections. While Cathy and Penny's general instruction practices were considered insufficient, they both included a variety of examples of instructions for data analysis. Like Sonia, both Cathy and Penny provided their students with the opportunity to experience multiple data sources, a best practice for the support of data analysis skills.

Finally, Margot's instructional practices were low in every section, providing evidence of what instruction that is need of improvement can look like. Most interesting, while Margot was lacking in general instruction, she excelled in her attempts to encourage data analysis skills, receiving the highest score for data analysis instruction in this study. Not only did Margot provide her students with the opportunity to work with multiple data types, but she was also the only teacher in this study who provided evidence of an attempt to integrate math and science.

Quantitative and qualitative analyses of the TIPs show that while no teacher in this study provided the researchers with exemplar TIPs for both categories of general instructional practice and data analysis instruction, neither was any teacher deficient in both. Teachers who lacked general instruction practices were strong in their incorporation of data analysis. Conversely, teachers with the best general instruction practices likely failed to incorporate instructional practices for data analysis. This suggests to the researchers that general best practice instruction may not be enough to encourage proficient data analysis skills. Instead, the quality of data analysis instruction is dependent on a combination of general best practice and data analysis specific instruction.

This data set furthermore provides an excellent example of diversity in quality of instruction being provided to students, both generally and with specific regard to data analysis instruction. Although the majority of teachers in this study utilized techniques to support student learning, these same techniques were not being as actively applied to the instruction of data analysis skills. As data analysis is a necessary skill to help our students be successful in life, and can be an effective means of helping to close our current gaps in achievement, we must find a means to help instructors see the value of data analysis instruction, but moreover a means with which to effectively instruct for data analysis (Silva, 2009; Zohar and Dori, 2003).

## **Trends in Student Work Samples**

The second question asked by this research study investigated the identification of trends in students' ability to analyze data as evidenced by their performance in their SWSs. The SWSs were collected as a class set, randomly sampled, and scored against the Oregon Department of Education (ODE) Science Inquiry Scoring Guide (ODE, 2011). In addition to this quantitative analysis, further qualitative analyses were conducted to specifically investigate student's ability to analyze results, draw conclusions, and communicate results.

Quantitative Trends in SWS. As seen in Figure 2, less than a quarter of the students in this study received high score on the data analyses section of their SWS. Not one student in this study received a score of 6, and only two received a score of 5 (both were students in Abigail's IB Biology class). Germann and Aram (1996) found that the greatest areas of student struggle when analyzing data are analyzing data, drawing conclusions, and communicating results. The quantitative scores students received for the SWSs were a summative score of the three categories of the ODE Science Inquiry Scoring Guide: Analyzing and Interpreting Results, which are also analyzing data, drawing conclusions, and communicating results (ODE, 2011). Each of these three categories is aligned with a bullet on the scoring guide, but this research study did not specifically investigate students' performance on each of these individual bullets. Instead, how students' performed with regard to data analysis specific skills is more



deeply investigating through the use of the qualitative coding scheme, which was specifically aligned with the analyzing data, drawing conclusions, and communicating results categories.

Qualitative Trends in SWS. While assessing the student work with the ODE Scoring Guide, the researchers simultaneously recorded qualitative observations with regard to student's ability to analyze results, draw conclusions, and communicate results. The three main codes of this scheme are aligned with the bullets of the ODE Science Inquiry Scoring Guide: Analyzing and Interpreting Results, as are the specific evaluation criteria categories Conclusions, Limitations of Data, and Content Knowledge.

*Analyzing Results: Data Agreement.* Aligned with the findings of Germann and Aram (1996), this study found that the greatest areas of student struggle were in the Analyzing Results domain. The majority of SWSs scored a zero for no evidence in the category of Data Agreement. For those students who did include a discussion of Data Agreement, the response generally focused on P-value and significance.

Many students struggled in determining the statistical significance of their results, and those students who found insignificant results often used language to try and make their results appear significant. An example of a Level 1 understanding or ability in Analyzing Results-Data Agreement which ignored an insignificant P-value was found in Penny's class, with:

*"The p-value I got was .4962 meaning that there was about a 50 percent chance that the plant would grow more with the solution or the control."*

This student went on to conclude that, based on the high P-value, it was "50-50 chance that the solution worked" and then discussed the results as if they were significant, incorrectly assuming that a 50-50 chance P-value would be considered significant. The student work in this study clearly aligned with the findings of Kanari and Millar (2004) in that very few students were comfortable with non-covariant variable relationships, most students ignoring their insignificant P-values and thus drawing inaccurate conclusions.

There were a few exceptional student responses in this study with regard to understanding or ability in Analyzing Results-Data Agreement. An example where the results did not support the hypothesis found in the work of a student from Margot's class when a student said,

*"Experimentally, this experiment was not as simple. We conducted the experiment expecting a clear pattern to develop. In the end, we saw no such pattern."*

This student accepted their non-covariant data, and was able to successfully discuss why the data showed that there was no relationship between the independent and dependent variables (Kanari and Millar, 2004). The ability to discern between non-covariant and covariant data, as discussed in Kanari and Millar (2004), is of great value in encouraging higher order cognitive skills for the future thinkers of our global

community. Some of the largest issues facing our planet, such as global climate change and world hunger, do not fit comfortably into linear graphs. Thus encouraging students to think “outside the box,” to experience a variety of data sources and types, helps to promote higher order thinking necessary for scientific literacy later in life (Hug and McNeill, 2008; Kanari and Millar, 2004).

*Analyzing Results: Data Source.* Similarly, half of the SWSs received a score of zero for no evidence in the category of Data Source. As Data Source was not listed on the ODE Science Inquiry Scoring Guide and many teachers did not specifically request students to include this information in their conclusions, it makes sense that many students failed to include this in their discussions. As Hug and McNeill (2008) found in their research, students acknowledging themselves as the source of their data not only encourages more effective data collection, it also contributes to richer discussions of limitations.

An example of a Level 1 understanding or ability in Analyzing Results: Data Source was found in Valerie’s class when her student wrote,

*“My data had some human error on the plants.”*

This student vaguely identifies the data as his or her own with the use of the word “my,” but does not discuss the “human error” as his or her own. An example of a Level 2 understanding or ability in Analyzing Results Data Source, a student in Felicia’s class wrote,

*"In the experiment that my group and I did, we were testing how the surface area affects the reaction rate of Alka-Seltzer."*

This student takes direct ownership of the experiment and the results, as they were produced by his or her group. According to the findings of Hug and McNeill (2008), students who identify themselves as the source of their data are often better able to explain their experimental processes as well as discuss the limitations of their data.

*Drawing Conclusions: Conclusions.* The category of Drawing Conclusions: Conclusions was specifically listed on both the ODE Science Inquiry Scoring Guide (ODE, 2011). Because this category, as well as the categories Drawing Conclusions: Limitations of Data and Communication Results: Content Knowledge, are listed as required items on the ODE Science Inquiry Scoring Guide, the teachers in this study were specifically made aware of these items when they were presented with copies of the ODE Science Inquiry Scoring Guide during their professional development training. As a result, this research found that evidence of the Drawing Conclusions: Conclusions category was present in at least a vague or unclear manner in every SWS. As an example of Level 1 understanding and ability in Drawing Conclusions: Conclusions, a student in Margot's class wrote,

*"[Providing three sets of averages] This shows that the distance did not significantly slow the ball bearing and explains why the distances traveled were relatively close."*

This student's conclusion statement is vague and does not relate to the hypothesis. A conclusion, as a scientific explanation, should include a statement of claim, evidence,

and reasoning (Ruiz-Primo et al., 2010). The majority of students in this study failed to include one or more essential element in their conclusions. This agrees with other findings in this study that components of data analyses that are not explicitly mentioned on the ODE rubric were the most likely to be ignored in the SWSs.

As an example of Level 2 understanding and ability in Drawing Conclusions: Conclusions, a student in Joe's class provided specific numerical data, similar to the Level 1 Drawing Conclusions: Conclusions student above, but brought the information together when he or she wrote,

*"From the data collected, it can be concluded that the hypothesis was correct. The cold tubes took longer to break up the Alka-Seltzer then the hot tubes. Heat speeds up the reaction because everything is move faster so there are more collisions between the atoms." [sic]*

This student made a claim specifically tied back to the experimental results, and stated that the results verify the hypothesis. The student used evidence from the experiment to justify the claim and explained why the results make sense using scientific reasoning.

*Drawing Conclusion: Limitations of Data.* As with Drawing Conclusions: Conclusions, the category of Drawing Conclusions: Limitations of Data was also specifically listed on both the ODE Science Inquiry Scoring Guide and Science Content Standards (ODE, 2011). Again, evidence of this category was present in at least a vague or unclear manner in every SWS (Table 7), as every work sample prompt specifically

asked for its inclusion of limitations. An example of Level 1 understanding and ability in Drawing Conclusions: Limitations of Data in Felicia's classroom was,

*"Because we did not start it [the experiment] in time it could have made reaction rate shorter then is actually was."*

This student provided minimal evidence that the design, procedures, and data had been reviewed to identify sources of uncertainties. On the other hand, an example of Level 2 understanding and ability in Drawing Conclusions: Limitations of Data was found in Sonia's class, when her student wrote:

*"Some limitations observed in this lab were an inconsistent method of measuring pulse. This would affect the lab because it causes the subject's health the chance to recoop [sic] from the effort therefore changing the difference in heartbeat. If there was a machine that measured the heart rates accurately the change in heartbeats of the subject would be more reasonable."*

As designated by the ODE Science Inquiry Scoring Guide and Science Content Standards (ODE, 2011), this student provided sufficient evidence that the design, procedures, and data had been reviewed to identify sources of uncertainties and further discussed how these limitations of the data may have affected the results. Furthermore, this student is utilizing the same higher order thinking skills discussed by Hug and McNeill (2008) to apply their understanding of their experimental results to real world barriers and limitations.

*Communicating Results: Patterns/Inferences.* Finally, in Communicating Results, the majority of students received a score of zero, showing signs of struggle with the identification of patterns or the development of inferences. An example of Level 1 understanding and ability in Communicating Results: Patterns/Inferences was found in Abigail's classroom with,

*"Therefore the mass of the celery would never reach zero even if the water was all gone, and this was why this trend was seen."*

This student vaguely infers that a trend was seen because the mass of a stick of celery will never be zero, but fails to further explain how this is related to any patterns he or she may have observed. This conceptual transition, of transferring direct observations to identifying relationships, is a common challenge for many students (McDermott, Rosenquist, and van Zee, 1987; Shah and Hoeffner, 2002; Shauble, Glaser, Duschl, Schulze, and John, 1995).

Examples of Level 2 understanding and ability in Communicating Results: Patterns/Inferences were also present in this data set. With regard to Patterns, a student in Joe's class wrote,

*"With these findings, it can be seen that the time necessary to complete the reaction will increase given that the temperature decreases."*

This student identifies a relationship in their data, linking time and temperature. Another example of Level 2 understanding and ability in Communicating Results:

Patterns/Inferences, this time to an inference with regard to experimental design, was in Sonia's class when her student wrote,

*"The large standard deviation implies that the accuracy of our data was bad."*

This student understood that a large standard deviation could be a sign of erroneous data and was able to infer that such results should cause one to question the accuracy of their results.

Finally, an excellent example of Level 2 understanding and ability in Communicating Results: Patterns/Inferences, was also from Sonia's class, when her student wrote:

*"Another interesting dynamic to consider is that the basketball players were currently in season at the time of this data collection. An athlete that is currently in season for any sport is obviously going to be in better shape than an athlete who has had 2 to 3 months off heavy physical activity. It is possible that this difference added to the intensely different average difference in bpm values."*

This student's comment illustrates the pattern that athletes had significantly different heart rates than non-athletes. This student then went on to infer that, from this pattern, a reason might have to do with their level of physical activity. Sonia and Joe's TIPs both received high scores, and the science inquiry activities that they assigned were consider open-ended. These findings come in agreement with Hofstein et al. (2004)



who found that students who participated in open-ended inquiries were better able to make higher-level cognitive connections. Further, these results suggest that as the complexity of an experiment or challenge increases, so does the need for students' ability to make inferences and identify patterns in order to make complete and appropriate sense of their experimental results (Hug and McNeill, 2008).

*Communicating Results: Content Knowledge.* The qualitative area of investigation with the least amount of student struggle was Communicating Results: Content Knowledge, which again was listed as a required category on the ODE Science Inquiry Scoring Guide and in the Science Content Standards (ODE, 2011). An example of Level 1 understanding and ability in Communicating Results: Content Knowledge was found in Penny's class when her student wrote,

*"Sugar is like nitrogen because it helps plants grow on its own, but when mixed with all of the other elements of fertilizer, it would help even more."*

This student attempted to bring in some outside knowledge, but their use of the content was inaccurate. For a student to effectively make an argument, they must have both the ability to analyze their data and the accurate content knowledge necessary to explain their understanding scientifically (Hug and McNeill, 2008; Windschitl et al., 2008, Ruiz-Primo et al., 2010).

An example of Level 2 understanding and ability in Communicating Results: Content Knowledge was the appropriate and correct use of the velocity equation ( $velocity = \frac{distance}{time}$ ) throughout Margot's class set of SWSs. The majority of Margot's students were able to appropriately apply their knowledge of the velocity equation to correctly predict how far a projectile would travel. As Margot's TIP was the lowest scoring in this study, the researchers question whether it was this integration of math and science that helped students to construct content knowledge strong data analyses. The integration of math and science has the ability to not only enrich a student's experience of both subjects, but can also help provide a strong foundational understanding with greater longevity then if the two subjects had been taught independently (Keiler, 2007).

Trends in Student's Data Analysis Skills Summary. Overall, the quantitative and qualitative analyses of the SWSs revealed that there are many areas of student struggle with regard to data analysis. A common trend was in the effect of templates on student's ability to draw conclusions. In agreement with the findings of Ruiz-Primo et al. (2012), teachers who do not provide enough scaffolding may not have met their students' needs for gaining access to higher order cognition. Likewise, providing too much scaffolding may have prevented students from being able to express their higher order thinking. In this study, many students were provided templates as a means of scaffolding. The templates used by the teachers in this study were either obtained, or

based off templates obtained, from the ODE website. Students who used these scaffolded templates often provided SWSs that lacked the “cohesive” quality required to receive a 5/6 on the ODE Science Inquiry Scoring Guide (ODE, 2011). These findings agree with the findings of Ruiz-Primo et al. (2010). Scaffolding is an essential ingredient for ensuring student success, but that further investigation is necessary into what the right amount of guidance is and how to properly time the release of this guidance.

Another trend was in students’ desire to have “correct” or “significant” results, thus negatively affecting their ability to properly analyze their results. These findings agree with those found by Kanari and Millar (2004), where students would go so far as to falsify their results in order to find correlations where none existed. Moreover, these findings suggest that students may lack an understanding of the nature of science. If students see science as the pursuit of proof or based in the desire for verification, they are leaving our classrooms with an inaccurate view of the goal science. Nature of science instruction is multifaceted, but should include an understanding of expectations in experimentation. Although an experiment should be organized in such a way so as to test a hypothesis, the results are dependent on the data and should not be influenced by a researcher’s desire for accuracy (Lederman, Abd-El-Khalick, Bell, and Schwartz, 2002).

Students’ also consistently failed to include content knowledge in their analyses. Quite often content knowledge was contained in the Background Research and failed to be included in the Data Analysis section of the SWS. If students are not using their

content knowledge to explain their experimental results, then what is the purpose of the experiment? Or vice versa, what is the purpose of the content? This disconnection of content and cognition presents an important predicament for science education researchers to ponder in the future. Although content knowledge and cognitive skills can be taught independently, successful integration of these two facilities yields higher quality scientific explanations (Hug and McNeill, 2008; Ruiz-Primo et al., 2010)

Finally, the ODE Science Inquiry Scoring Guide: Analyzing and Interpreting Data Rubric does not include the all of the subtopics included in this study. While the Science Inquiry Scoring Guide does have three main categories aligned with this study (Analyzing Results, Drawing Conclusions, and Communicating Results), not all of the evaluation criteria in this study are specifically listed on the rubric. While the evaluation criteria alone can vary in the level of higher order cognition being used to support each descriptor, data source and inferences are not listed on the ODE Science Scoring Guide. As such, it is currently up to the teacher or students to assert the value of these data analysis components.

## Case Study Comparisons

The final question of this research study examined the relationships between instructional practices and students' ability to analyze data. The researchers originally predicted that lower scoring TIPs would be associated with lower scoring SWSs, and likewise that higher scoring TIPs would be associated with higher scoring SWSs. In several cases, this was not what was found.

Low TIP-Low SWS: Cathy-Margot-Penny. Cathy and Penny each received a summative score of 7 on the TIP, and Margot received the lowest score in the study, a summative score of 6. Each teacher had unique struggles. Cathy scored high for her implementation of student-centered classroom roles, but scored as low performing in content and cognitive skills and in assessment for learning. The students in Cathy's classroom had a sum score of 23. Penny scored low on classroom roles and assessment for learning, but received high score for balanced content and cognitive skills. Penny's classroom had a sum score of 23. Margot scored as low performing in every section of the TIP, and her classroom had a sum score of 19. All of these SWSs in these classrooms received an overall class low performing average of 3.

An interesting observation is that of the cases studied in this research project, these three classrooms conducted the shortest units of instruction. Cathy and Penny conducted units that ran for only 17 days and Margot's unit was 23 days. As multiple exposures to the science inquiry cycle promotes students' higher order cognitive skills

(Germann and Aram, 1996; Hofstein et al., 2004), these short units of instruction may not have allowed the students enough time to practice their data analysis skills or experience multiple iterations of the science inquiry cycle, thus resulting in low scoring student work.

Another commonality is that each teacher provided a template to the students in their classroom. The ecology unit teachers, Cathy and Penny, both provided mandatory templates for their turnip seed science inquiries, while Margot provided an optional template to her physics class that many students utilized for their projectile inquiries. These templates, meant to be used as scaffolds, restricted students ability to express their understanding. The templates had limited space for students to construct their responses, usually varying between three to seven blank lines. Breaking student responses into blocks of comprehension essentially turned the scientific inquiry experience into an exercise of filling-in-the-blanks. These over-scaffolded science inquiries were unable to be used as effective performance-based instrument, as they no longer represented the capabilities of the students who completed the templates. When highly scaffolded templates were used for science inquiry assessment of data analysis skills, the students completing them were incapable of achieving the cohesive and in depth response required for a level 5/6 response on the ODE Science Inquiry Scoring Guide: Analyzing and Interpreting Data Rubric (ODE, 2011).

Finally, every classroom that scored low on both the TIP and the SWS had 26 or more students. This finding agrees with Fowler and Walberg (1991) and Fowler (1995),

who found correlations between class size and student achievement. Both investigations found that class size has a direct relationship with student achievement. Zohar and Dori (2003) found that student achievement of higher order thinking skills in science were linked to participation in scientific inquiry, and that lower achieving students actually benefited more from participation in science inquiry investigations than their higher achieving counterparts. Thus, with large class sizes these teachers may not have been able to help their students succeed in attaining the higher order thinking skills necessary to effectively analyze their data.

High TIP-Low SWS: Felicia-Sonia. Two of the cases included in this study scored relatively high with regard to their TIPs and yet their SWSs still scored relatively low. Felicia and Sonia received the highest TIP scores in this study, receiving high performing scores in all three sections of the TIP Scoring Guide. The SWSs from Felicia's class averaged low performing scores of 3. Similarly, Sonia's IB biology class received low performing scores on their SWSs even though Sonia's TIP scored high performingly in two categories out of three. Sonia scored as high performing in her balance of Content and Cognitive Skills and in her Assessment for Learning, but received low score for her Classroom Roles, which tended toward teacher-centered instruction, with students as passive participants.

Both classes had 16-25 students and implemented longer units, Felicia's chemical reactions unit spanning 5 weeks and Sonia's anatomy and physiology unit

spanning 28 days. Both units were relatively comprehensive, covering at least three major content learning goals in addition to the science inquiry skill goals. Another commonality between these units was that each provided an optional template for students to use while conducting their inquiries. The students who chose to use the templates generally scored lower than those who chose not to use them, and the majority of students in both classes opted to use the templates. The use of the scaffold may have hindered students in their ability to successfully achieve a higher score due to their lack of cognitive cohesion required of a highly proficient score on the ODE Scoring Guide.

A final observation was that while Felicia had the highest scoring TIP in the study, her classroom also had the highest number of students classified as English Language Learners (ELL) (11-30% of students). Although studies have found that the achievement gap between ELL students and non-ELL students is narrowing, there is still a significant disparity between the two groups (Kao and Thompson, 2003; Snow and Biancarosa, 2003). Science inquiry has been found to be an effective means of increasing higher order thinking skills among lower achieving students (Zohar and Dori, 2003; Lee, 2005). ELL students can gain great benefit from scaffolded instruction. As the use of the template was the only scaffolding instructional technique seen in either of these TIPs, this again leads to the question of whether the use of a template is an effective means of scaffolding instruction.



In summary, this research did not find conclusive data as to why these high performing TIPs resulted in lower performing SWS. The data sources utilized in this study were unable to provide evidence for a clear interpretation, and the inclusion of additional data sources would greatly benefit this type of research in the future.

High TIP-High SWS: Abigail-Joe. Two of the classrooms in this case study received high performing scores for both their TIPs and the SWSs they produced. Both Abigail and Joe scored high on their TIPs, and the SWSs also scored relatively high against the ODE Science Inquiry Scoring Guide. Like Abigail, Joe received scores of high performing for every category in his TIP. Similarly, his students received an average score of high performing. Abigail scored as high performing in her balance of Classroom Roles and in her Assessment for Learning, but received low score for her imbalance of Content and Cognitive Skills, which provided incomplete and inconsistent opportunities for students to construct meaning, apply content, and practice higher-order cognitive skills.

Interestingly, Joe provided his chemistry students with a template for their Alka-Seltzer inquiries, while Abigail did not provide any such template to her cell biology students for their osmosis inquiries. Unlike the other classes in this study, Joe's students were the most likely to utilize all of the space provided to them to answer the questions. There are not written instructions on the template specifically telling the students to use all of the space provided to completely answer the prompt questions, but Joe's class was the only to do so. This raises the question whether Joe may have

prompted his students with undocumented instructions, or whether another factor prompted students to give full and complete responses. Furthermore, while Abigail did not provide a specific template for student responses, she did provide students with a copy of the ODE Science Inquiry Scoring Guide, thus alerting them to the expectations for SWSs. These findings agree with the suggestions of Ruiz-Primo et al. (2010), which propose that there may be a balance in finding the ideal level of guidance to best assist students in gaining understanding. Students need some guidance, helping them to understand what expectations they are being held accountable to, but also providing them with the freedom to express themselves and “think outside the box,” particularly with regard to higher order cognitive skills.

Case Study Summary. An interesting result of this study was that, while the lowest TIP scores was associated with a low SWS score; the highest TIP scores were not necessarily associated with high SWS scores. High TIP scores were found to be associated with either low or high SWS scores, but low TIP scores were not associated with high-scoring SWSs. No relationship was found relating student ability to teachers’ years of experience, teachers’ level of education, nor student demographic indicators of sex or ethnicity. A possible relationship was observed between student achievement in data analysis and predominance of a template in the science inquiries. Each of these findings is in agreement with documented research. The findings of Ruiz-Primo et al. (2010) suggest that templates can be an effective means of scaffolding instruction, but

should be used with caution as over use can in fact be detrimental to student learning. Interestingly, although data analysis instruction was found to be relatively uniform across the classes participating in this study, the resultant quality of student work was not. As such, specific instruction in data analysis did not predict better student performance in data analysis. The sensitivity of the coding of the TIP qualitative coding scheme will be further discussed as a limitation of this study.

### **Discussion Summary**

After analyzing the TIPs and the SWSs separately, relationships between the two were examined. High scoring TIPs were related to either high or low scoring SWSs, but low scoring TIPs only related to low scoring SWSs. This appears to imply that while best practices for instruction can result in higher levels of student achievement, instruction neglectful of best practices appears to only result in low scoring student work. Furthermore, effective instruction does not necessarily mediate the score a student will receive on an assessment. A variety of factors can influence a student's achievement academically, including teacher understanding of the nature of science (Lederman, 1998), student prior knowledge (Orion and Hofstein, 2006), and student perception of science (Gibson and Chase, 2002).

An analysis of teacher instructional portfolios found that many teachers in this study were utilizing best practices in teaching. With regard to Classroom Roles, more than half of the teachers in this study were conducting their instruction in a manner that

promoted student-centered learning and active student participation in the classroom. The content and cognitive skills were well-balanced in more than half of the TIPs, and students were being provided multiple opportunities to practice both, often at the same time. The Assessment for Learning practices implemented by more than half of the teachers were found to be developmentally appropriate, frequently utilized, and allowed for student self-reflection. The teachers participating in this study self-selected to participate due to their interest in professional development. As such, it could be assumed that the results of this study may over-represent positive general instructional practices.

Qualitative investigations into the TIPs as a holistic scoring guide revealed that direct instruction was utilized in every classroom. Similarly, formative assessments were used in every classroom, but no evidence of their use to alter instruction could be ascertained with this instrument. Varying levels of scaffolding were used in all but one classroom, most often in the format of providing students with a template with which to construct their scientific inquiries. All teachers but one (Abigail) recorded data-analysis-specific instruction in their TIPs. Data analysis instruction was often delivered via direct instruction and scaffolds in the form of templates. No formative assessments of students' data analysis skills were observed in this research study. Furthermore, while Abigail did not provide her IB Biology students with documented data analysis instruction, her students did well both quantitatively and qualitatively. This leads the researchers to question whether these students were previously instructed on data analysis prior to this unit, or if they possibly received undocumented instruction to

better prepare them for science inquiry investigations prior to their participation in this study.

The prevalent use of templates in these science inquiry units was quite informative. In an effort to scaffold science inquiry instruction for students, teachers provide such templates. The templates used in this study were either directly selected or formatted from the ODE Anchor Templates<sup>1</sup> which are available to teachers with the click of a button off of the ODE website. The ODE Anchor Template divides the Analyzing and Interpreting Results section into small distinct pieces: Results, Conclusion, Explanation, Design Review, and Extension. While each of these components is indeed necessary for a comprehensive analysis of data, breaking a discussion in this way does not allow the student an opportunity to formulate a cohesive narrative of their understanding. Moreover, the ODE Anchor Templates are made available to teachers, and the general public, without calibration to ensure their validity or reliability. While these templates may be helpful to students as means of scaffolding, research suggests that successful scaffolding begins with closed-ended templates that are slowly removed until as students become prepared to participate in open-ended inquiries (Hofstein et al., 2004; Ruiz-Primo et al, 2010).

Conversely, of the sixty-four SWSs scored, only fifteen were considered high performing. Students' greatest area of difficulty was in the domain of Analyzing Results, specifically with regard to understanding data agreement and identifying a data source. Many students struggled with their ability to construct conclusions and to identify the

---

<sup>1</sup> [www.ode.state.or.us/wma/teachlearn/testing/scoring/guides/2011-12/science\\_inquiry\\_notebooktemplate\\_hs.pdf](http://www.ode.state.or.us/wma/teachlearn/testing/scoring/guides/2011-12/science_inquiry_notebooktemplate_hs.pdf)

limitations of data. Finally, in Communicating Results, while many students brought content knowledge into their discussions, very few students were able to make inferences based on their data.

Overall, students appeared to score highest in the areas that were specifically identified by the ODE Science Inquiry Scoring Guide, such as the incorporation of content knowledge and the listing of experimental limitations. Student inclusions of these two evaluation criteria were most often vague and/or inaccurate. Overall, most students seemed unable to incorporate the identification of patterns or understanding of data agreement into their discussions, both of which are strong examples of higher order thinking skills (Hug and McNeill, 2008). As these expectations are not set in the ODE Science Inquiry Scoring Guide: Analyzing and Interpreting Results, it would make sense for the students who were aware of the scoring guide to not include them. Recently, the state of Oregon has decided to adopt the Next Generation Science Standards (NRC, 2012). This new set of standards is built from a three part framework emphasizing science and engineering practices, crosscutting concepts, and core ideas. The dimension of practices specifically emphasizes the higher order cognitive skills required to successfully participate in science inquiry. Like the ODE Science Inquiry Standards, the Next Generation Science Standards require students to draw conclusions, identify the limitations on their data sets, and draw up content knowledge in their scientific explanations. Unlike the ODE Science Inquiry Standards, the Next Generation Science Standards further require students to investigate data agreement via statistical

analyses, identify and analyze the source of the data, and present the results in such a way as to highlight patterns and relationships (NRC, 2012).

### **Limitations of the Study**

This mixed-method, multiple-case study was designed to analyze a preexisting data set. A major disadvantage of this study was that it was designed to investigate whether any relationships exist between teacher instructional practice and students' ability to analyze data, a question that was not originally intended by the primary authors. As the data sources used had been collected prior to the beginning of this study, they were not originally intended to be used for the examination of data analysis skills specifically. As a result, the analysis of the data sets was limited to the scope of the original study (Saxton, pre-publication). Furthermore, an extensive amount of time and energy was required for the researchers to acquaint themselves not only with the data sources, assessment instruments, and the process of data collection; but with the much larger research project in which this project was contained.

The teachers who chose to participate in this research study did so of their own volition. Thus, the teachers who participated in this study were those who were interested in participating in professional development. With this in mind, this study lacks a holistic representation of instructional practices happening across the local and national district levels. In future studies the researchers would strongly recommend

reaching out to a larger instructional community, particularly those teachers who do not normally participate in professional development.

It is also important to consider that the lack of detail the researchers have with regard to the students in this study. While access to classroom level demographics was made available by the districts involved, not student level demographics were provided. Most importantly, no information was available to the researchers regarding student prior academic experience. The researchers did not know what level of experience the students had with science inquiry prior to this study, nor their prior content knowledge. Future studies should request student level demographic data, including a pre-participation survey of previous experience and understanding.

Another issue encountered with this data set was the sensitivity of the literature-based TIP qualitative coding scheme. This schema used three levels of the assessment to code the TIPs, from 0 to 2, on a variety of evaluation criteria. This research found that the level 1 and level 2 lacked the sensitivity necessary to represent to diverse practices observed in the TIPs. This research could be improved in the future with the addition of assessment levels to the qualitative TIP coding scheme to help distinguish the differing levels of data analysis specific instruction.

Finally, obtaining IRR was a challenge with both quantitative scoring guides used in this study. The researchers were unable to reach IRR with the TIP scoring guide, which is currently under investigation for its validity and reliability (Saxton, pre-publication). For the purposes of this research, consensus scores were used for TIP analysis. The



researchers found that personal experience was a great factor in score selection, and that the individual raters often experienced difficulty maintaining intra-rater reliability. Further research studies IRR should be achieved to confirm the reliability of the TIP scores used in this study.

The ODE Scoring Guide, which has been adopted by the state of Oregon as the primary means of assessing student's data analysis abilities, also provided exceptional challenges for the researchers in the study to obtain IRR. While IRR with the ODE Scoring Guide: Data Analysis and Interpretation was achieved in this study, arriving at the point of agreement required multiple rounds of calibration. Overall, it took to approximately 15 hours of discussion (five 3-hour meetings) for the researchers in this study to reach IRR using the ODE Scoring Guide: Data Analysis and Interpretation. One confounding factor was the researchers own content knowledge, which was not always expert in the subject being analyzed. For example, not all researchers had a bachelors degree or higher in physics, but were attempting to score physics work samples. As this is the complete science inquiry scoring guide being used across the state, it is important to consider the implications of these challenges. If it takes four researchers 15 hours of training to calibrate themselves to reliably use this single section of the four part science inquiry scoring guide, what challenges are teachers currently facing while attempting to use the scoring guide in its entirety as an assessment instrument in their classrooms? Is this scoring guide being used effectively and reliably by the teacher population it is intended to serve? For this reason, further research should be done to investigate the

validity and reliability of the ODE Scoring Guide as an assessment tool should it be used for future research.

### **Questions for Further Research**

Probably the most significant finding of this research was the effect of templates of scaffolding on student achievement. The results of this research suggest that the overuse of templates may have in fact been detrimental to student success in creating a cohesive explanation of their analyses. This comes in agreement with the findings of Ruiz-Primo et al. (2010) and Simon and Klein (2007) who suggest a middle ground must be struck between under and over-scaffolding science inquiry assignments to best achieve student attainment of higher-order cognitive skills. The researchers strongly suggest further research be conducted to investigate the effect of scaffolding, particularly in regard to the use of templates as a means of scaffolding in science inquiry investigations.

Another question presented by this study was how to effectively assess teachers' formative assessment practices. The TIP is designed as a holistic means of assessing classroom instructional practices and aids teachers in designing units of instruction that are thoughtfully and thoroughly planned. The TIP also documents teacher's expectations for knowledge, skills, and experiences outcomes as well as their planned assessments and pedagogical reflections. The TIP does not record classroom

discussions, teacher to student interactions, or student to student interactions. An important finding of this study was that, although teachers were directly requested to document informal formative assessments in their TIPs, they rarely provided artifacts for analysis. Furthermore, adjustments or changes to instructional practices are only documented if the teacher participant so chooses to document them. The TIP assessment instrument is a newly developed, iterative assessment tool for investigating instructional practices and the TIP instrument used in this study was the pilot version of the TIP used in the first year of study. As such, this version of the TIP did not include the frequency of prompts or investigative probes currently being used promote teacher documentation of changes to instruction, particularly with regard to formative assessment. This research study would benefit from being conducted again with the latest form of the TIP instrument, as well as additional data sources for the documentation of informal formative assessment and classroom discourse.

Finally, this study specifically investigated how student's data analysis abilities relate to overall instructional practices, but analysis of the TIPs inspired further questions into the nature of the specific tasks being asked of the students. As not all science inquiry tasks are created equal, this study may have neglected to account for the variety of tasks being present to the students. Do students being asked lower order thinking questions have the ability to use higher order thinking skills? Are students up to the challenge of answering higher order prompts or tasks given that the majority of the students in this study submitted SWSs that were not proficient in data analysis? This

research suggests that the nature of the task may have affected students overall ability to perform both with regard to content and cognitive ability. This agrees with the findings of Shavelson, Baxter, and Gao (1993) who, during their investigation into the validity of performance based assessments, found that a student's performance depends on the requested task they are asked to perform. We must ask students to perform advanced investigations if we hope to receive work samples that reflect advanced understandings. The growing body of literature on how to best serve students in encouraging higher order cognitive skills would benefit from the inclusion of an investigation into the nature of the tasks being assigned in classrooms.

### **Implications for Instructional Practice**

This research further supports already existing evidence that data analysis skills are not necessarily intrinsic, and must be cultivated for students to be able to successfully interpret their results. The majority of students in this study produced work samples that were not proficient with regard to their ability to analyze data. To help encourage data analysis skills, teachers should provide students with multiple opportunities to experience science inquiry. The researchers suggest that teachers being the year with an assessment of their students' science inquiry skill set, and then use this information to scaffold to their student's needs accordingly. Furthermore, teachers must mindfully select means of scaffolding to best ensure student success. The researchers view the template scaffold as an excellent tool for encouraging the growth

and development of students' skills, but it must be used as a means of assistance and not a crutch. By the end of an academic year, students should have experienced enough science inquiry activities to allow for the removal of much of the scaffolding the template provides, allowing the student to experience a truly open-ended inquiry.

Another implication of this research is with regard to students' desire to get the "right answer." One of the most commonly identified areas of student struggle in this research was the frequency with which students ignored negative results or made incorrect statements with regard to their P-values in an attempt to make their results appear more substantive. According to Lederman, Abd-El-Khalick, Bell, and Schwartz. (2002), this common practice is directly related to students' understanding of the nature of science. If students are being taught that science inquiries have a right answer, this is a strong indicator that either the science inquiries are being implemented incorrectly or that students do not understand the nature of how scientific inquiries function. This is highly concerning, as scientists must be held to a high degree of impartiality with regard to their results, ensuring that data is not ignored or skewed in search of an ultimate goal. Thus, instructors should include concepts related to the nature of science in their instruction, explicitly citing the value of evidence based reasoning.

With regard to the challenges faced by the researchers in achieving IRR, the researchers strongly urge teachers to participate in regular calibration activities. These professional development trainings provide teachers with the opportunity to moderate both their intra-rater and inter-rater reliability when using state mandated scoring

guides. In order to ensure that these scoring guides are being effectively and reliably, grade level calibration experience is of the utmost importance.

Finally, this research found that no teachers explicitly incorporated data analysis best practices, and only four teachers vaguely incorporate the data analysis best practice instructional techniques. These findings emphasize the need for professional development in the educational community, particularly science teachers at the high school level. Teachers need an opportunity to not only collaborate on the development of curriculum, but the time and space to consider their beliefs as educators. Although most teachers will tell an investigator that they value higher order cognition (Zohar and Dori, 2003), many teachers are not instructing for these necessary life skills. This may be because they do not think the students are capable of higher order cognition, or it may be that their means of implementation are ineffective (William, Lee, Harrison, and Black, 2004; Zohar, 1999). Furthermore, the Next Generation Science Standards strongly emphasize science practices that rely on higher order cognition, so professional development on how to develop these skills is becoming necessary for job security. Ultimately, professional development should provide the opportunity for teachers to gain and build upon the instructional skills necessary to effectively teach for higher order thinking.

## References

- Anderson, R.D. (2002). Reforming science teaching: What research says about inquiry. *Journal of Science Teacher Education*, 13(1), 1-12.
- Andrade, H., Du, Y., and Wang, X. (2008). Putting rubrics to the test: The effect of a model, criteria generation, and rubric-referenced self-assessment on elementary school students' writing. *Educational Measurement: Issues and Practices* 27, (2), 3-13.
- Black, P. and Williams, D. (1998). *Inside the Black Box: Raising Standards through Classroom Assessment*, King's College School of Education, London.
- Dean, D.J. and Kuhn, D. (2006). Direct instruction vs. discovery: The long view. *Science Education*, 91(3), 384-397.
- Elby, A. (1999). Another reason that physics students learn by rote. *American Journal of Physics Teachers*, 67 (S1), S52-S57.
- Fowler, W. J. (1995). School Size and Student Outcomes. *Advances in Educational Productivity*, 5, 3-26.
- Fowler, W. J., and Walberg, H. J. (1991). School Size, Characteristics, and Outcomes. *Educational Evaluation and Policy Analysis*, 13(2). 189-202.
- Germann, P.J. and Aram, R.J. (1996). Student performances on the science processes of recording data, analyzing data, drawing conclusions, and providing evidence. *Journal of research in science teaching*, 33 (7), 773-798.
- Gibson, H.L. and Chase, C. (2002). Longitudinal impact of an inquiry-based science program on middle school students' attitudes toward science. *Science Education*, 86, 693-705.
- Higgins, B., Miller, M., and Wegmann, S. (2006). Teaching to the test... not! Balancing best practices and testing requirements in writing. *International Reading Association*, 60(4), 310-319.
- Hofstein, A., Navon, O., Kipnis, M. and Mamlok-Naaman, R. (2005). Developing students' ability to ask more and better questions resulting from inquiry-type chemistry laboratories. *Journal of Research in Science Teaching*, 42(7), 791-806.
- Hug, B. and McNeill, K.L. (2008). Use of First-hand and Second-hand Data in Science: Does data influence classroom conversation? *International Journal of Science Education*, 30 (13), 1725-1751.
- Kanari, Z. and Millar, R. (2004). Reasoning from data: How students collect and interpret data in science investigations. *Journal of Research in Science Teaching*, 7 (41), 748-769.
- Kao, G and Thompson, J.S. (2003). Racial and Ethnic Stratification in Educational Achievement and Attainment. *Annual Review of Sociology*, 29, 417-442.

- Keiler, L. S. (2007). Students' Explanations of their Data Handling: Implications for transfer of learning. *International Journal of Science Education*, 29(2), 151–172.
- King, F.J., Goodson, L., and Rohani, F. (1998). Higher order thinking skills: Definitions, strategies, assessment. *Assessment evaluation*. Tallahassee, FL. Center for Advancement of Learning and Assessment.
- Klahr, D., and Nigam, M. (2004). The equivalence of learning paths in early science instruction: Effects of direct instruction and discovery learning. *Psychological Science*, 15, 661 – 667
- Lederman, N. G., Abd-El-Khalick, F., Bell, R. L., and Schwartz, R. S. (2002). Views of Nature of Science Questionnaire (VNOS): Toward valid and meaningful assessment of learners' conceptions of nature of science. *Journal of Research in Science Teaching*, 39(6), 497–521.
- Lederman, N.G. (1999). Teachers' understanding of the nature of science and classroom practice: Factors that facilitate or impede the relationship. *Journal of Research in Science Teaching*, 36(8), 916-929.
- Linn, M. C., Lee, H.S., Tinker, R., Husic, F., and Chiu, J. (2006). Teaching and Assessing Knowledge Integration in Science. *Science*, 313, 1049-1050.
- Lovett, M. C., and Chang, N. C. (2007). Data-analysis skills: What and how are students learning? In M. Lovett, and P. Shah (Eds.) *Thinking with Data*. New York: Lawrence Erlbaum.
- Lomax, R., Maxwell West, M., Harmon, M.C., Viator, K.A, and Madaus, G.F. (1995). The Impact of Mandated Standardized Testing on Minority Student. *Journal of Negro Education*, 64 (2), 171-185.
- McDermott, L., Rosenquist, M. L. and van Zee, E. H. (1987). Student difficulties in connecting graphs and physics: Examples from kinematics. *American Journal of Physics* 55, 503-513.
- Miner, D.D., Levy, A.J., Century, J. (2009). Inquiry-Based Science Instruction- what is it and does it matter? Results from a research synthesis years 1984 to 2002. *Journal of Research in Science Teaching*, 47(4), 474-496.
- National Research Council. (1987). Education and Learning to Think. *Committee on Research in Mathematics, Science, and Technology Education*. Washington, D.C.: The National Academies Press.
- National Research Council. (1996). *National Science Education Standards*. Washington, DC: The National Academies Press.
- National Research Council. (2012). A framework for K-12 science education: Practices, crosscutting concepts and core ideas. *Next Generation Science Standards: Appendix F*. Washington, D.C.: The National Academies Press.



- Okhee, L. (2005). Science education with English language learners: Synthesis and research agenda. *Review of educational research*, 7(4), 491-530.
- Onwuegbuzie, A. J., and Leech, N. L. (2007). Sampling designs in qualitative research: Making the sampling process more public. *The Qualitative Report*, 12(2), 238-254.
- Oregon Department of Education. (2011). *2011-2012 Official Scientific Inquiry/Engineering Design Scoring Guide: High School*. Salem, Oregon. Office of Assessment and Information Services.
- Orion, N. and Hofstein, A. (1994), Factors that influence learning during a scientific field trip in a natural environment. *Journal of Research in Science Teaching*, 31(1), 1097–1119
- Popham, W. J. (1999). Where Large Scale Educational Assessment Is Heading and Why It Shouldn't. *Educational Measurement: Issues and Practice*, 18(3), 13–17.
- Ruiz-Primo, M. A., Li, M., Tsai, S.-P. and Schneider, J. (2010). Testing one premise of scientific inquiry in science classrooms: Examining students' scientific explanations and student learning. *Journal of Research in Science Teaching*, 47(5), 583–608.
- Ruiz-Primo, M.A. (2007). Exploring Teachers' Informal Formative Assessment Practices and Students' Understanding in the Context of Scientific Inquiry. *Journal of research in science teaching*, 44(1), 57-84.
- Sadler, R. (1989). Formative assessment and the design of instructional systems, *Instructional Science*, 18, 119-144.
- Schauble, L., Glaser, R., Duschl, R. A., Schulze, S., and John, J. (1995). Students' understanding of the objectives and procedures of experimentation in the science classroom. *The Journal of Learning Sciences*, 4(2), 131-166.
- Schoenfeld, A. (1988). When Good Teaching Leads to Bad Results: The Disasters of 'Well-Taught' Mathematics Courses *Educational Psychology*, 23(2), 145-166.
- Shah, Priti; Hoeffner, James; (2002). Review of Graph Comprehension Research: Implications for Instruction. *Educational Psychology Review*, 14 (1), 47-69.
- Shavelson, R. J., Baxter, G. E, and Gao, X. (1993). Sampling variability of performance assessments. *Journal of Educational Measurement*, 30(3), 215-232.
- Simons, K. and Klein, J. (2007). The impact of scaffolding and student achievement levels in a problem-based learning environment. *Instructional Science*, 35(1),41–72.
- Silva, E. (2009). Measuring Skills for 21st-Century Learning. *Phi Delta Kappan*, 90(9), 630-634.

- Snow, C., and Biancarosa, G. (2003). Adolescent literacy and the achievement gap: What do we know and where do we go from here? New York: Carnegie Corporation of New York.
- Stern,, L. and Ahlgren, A. (2002). Analysis of students' assessments in middle school curriculum materials: Aiming precisely at benchmarks and standards. *Journal of Research in Science Teaching*, 39(9), 889–910.
- Webb, M., and Jones, J. (2009). Exploring tensions in developing assessment for learning. *Assessment in Education: Principles, Policy and Practice*, 16(2), 165–184.
- Wiggins, G. (2011). A True Test: Toward more authentic and equitable assessment. *Phi Delta Kappan*, 92(7), 81-93.
- Wiggins, G. and McTighe, J. (2008). Put understanding first. *Educational Leadership*, 65(8), 36-41.
- William, D., Lee, C., Harrison, C. and Black, P. (2004) Teachers developing assessment for learning: impact on student achievement. *Assessment in Education: Principles, Policy and Practice*, 11(1), 49-65.
- Windschitl, M., Thompson, J., and Braaten, M. (2008). Beyond the scientific method: Model-based inquiry as a new paradigm of preference for school science investigations. *Science Education*, 92(5), 941-969.
- Zohar, A. and Dori, Y. (2003). Higher order thinking skills and low achieving students: Are they mutually exclusive? *Journal of the Learning Sciences*, 12(2), 145-182.
- Zohar, A. (1999). Teachers' metacognitive knowledge and instruction of higher order thinking. *Teaching and Teachers' Education*, 15, 413–429.

## Appendix A

### **Knowledge, Skills, Experiences Outcomes**

- 1) Insert the Knowledge, Skills, and Experiences Outcomes from your preliminary unit plan into the available sleeve(s) so you can compare it to the actual unit that was taught.
- 2) During your unit, please follow the Instructions below for this section of your instructional portfolio.

Use the 3 ring binder and the provided pockets to document artifacts from your unit that represent the knowledge, skills, and experiences outcomes for your unit.

Artifacts you might select for your portfolio include:

- **A revised knowledge, skills, and experiences outcomes page (blank copies provided), based on what actually happened in the classroom**
- Written plans of lessons (if generated ahead of time)
- Powerpoint slides or lecture notes
- Notes, problems, or questions you wrote on the white board or transparencies
- Description of any classroom activities relevant to the outcomes
- Notes written to yourself about the lesson
- Observation notes written to yourself about your students
- Handouts given that relate to the knowledge outcomes (ex. Notes, worksheets, laboratory instructions, problem descriptions, etc.)

## Assessments

- 1) Insert the assessments page from your preliminary unit plan into the back of this sleeve so you can compare it to the actual unit that was taught.
- 2) Please follow the Instructions below for this section of your Instructional portfolio.

Use the 3 ring binder and the provided pockets to document the assessments used for your unit – be mindful of both your formal and informal (questioning strategies, quick homework checks at the beginning of class, etc.) assessment practices and try to document both.

Artifacts you might select for your portfolio include:

- **A revised assessment page (blank copies provided), based on what actually happened in the classroom**
- **Prompts given to generate the Science Inquiry (SI)/Engineering Design (ED) work sample**
- Formative assessments given at the end of, or during, a class period to inform you of student progress
- Notes about questioning strategies used in your class (what questions were asked? How did you determine which student responses to accept? What information did your students' answers provide? How was that information used?)
- Homework
- Journal entries
- Summative assessments (unit exams, projects, etc.)
- Observation notes written to yourself about your students
- Rubrics used to score your students' assessments

**\*\*\*Please note: that full class sets of student SI/ED work samples will also be collected, but in this portfolio section you can select low, medium, and high examples of student work to illustrate your assessment practices.\*\*\***

### **Pedagogical Strategies**

- 1) Insert the pedagogical strategies page from your preliminary unit plan into the back of this sleeve so you can compare it to the actual unit that was taught.
- 2) Please follow the Instructions below for this section of your Instructional portfolio.

Use the 3 ring binder and the provided pockets to document the pedagogical strategies page used for your unit.

Please make sure you include:

- **A revised pedagogical strategies page (blank copies provided), based on what actually happened in the classroom**
- **Reflections about the pedagogical strategies**
- Any artifacts that demonstrate pedagogical strategies that were used (i.e. handouts you give students about group work, notes from class discussions, etc).

***\*\*You may not have many artifacts for this section, but your reflections are very important.***

***Please make sure to fill out the attached reflection prompts at the end of the unit\*\****

### **Pedagogical Strategies: Reflections**

There are many different pedagogical strategies and more than strategy can be used in any given lesson.

- 1) Of the pedagogical strategies used in your unit, please estimate the percentage of class time spent using each strategy for the unit as a whole.

- 2) Of the strategies used, which strategies were most effective in helping students better understand the content? Why?

Of the strategies used, which strategies were most effective in helping students develop a better understanding of science inquiry/engineering design?

- 3) Of the strategies used, which strategies were most important in helping your students develop their science inquiry/engineering design work sample?

## Appendix B

Teacher Instructional Portfolio (TIP) Rubric			
Score	Classroom Roles	Content and Cognitive Skills	Score
4	<ul style="list-style-type: none"> <li>Lessons and activities require students to primarily rely on their observations, reasoning and interactions with their peers to determine the accuracy of their results; the teacher facilitates student learning by pressing for higher cognitive levels of student thinking when needed.</li> <li>Lessons and activities require students to engage in inquiry (e.g., asking questions, making observations, collecting data, drawing conclusions) and problem solving (e.g., looking for patterns, making conjectures, justifying those conjectures) as a means of gaining deep conceptual understanding.</li> <li>Activities/investigations are cognitively challenging with students as active participants engaged in learning that is, in some respect, of their own design.</li> </ul>	<ul style="list-style-type: none"> <li>The unit is designed to target both content and higher-order cognitive skills, frequently at the same time, presenting students with opportunities to use both content knowledge and cognitive skills in the same activities, lessons, or investigation(s).</li> <li>Lessons, activities, and investigation(s) in the unit, when taken together, represent multiple and diverse varieties of opportunities for students to construct meaning, apply content, and practice higher-order cognitive skills.</li> </ul>	4
3	<ul style="list-style-type: none"> <li>Lessons and activities require students mainly rely on their observations and reasoning to determine the accuracy of their results; teachers occasionally take the lead to facilitate learning by posing or answering questions or focusing attention on an important concept or a contradiction.</li> <li>Lessons and activities require students to participate in inquiry and problem solving to explore concepts, after having completed tasks/activities the teacher structures to provide scaffolding of concepts and/or skills.</li> <li>Activities/investigations are staged so that students are mostly active participants contributing their own ideas, applying skills previously learned to a new context, or making connection to concepts previously learned.</li> </ul>	<ul style="list-style-type: none"> <li>The unit is designed to target both content and higher-order cognitive skills with frequent opportunities for students to use both content knowledge and cognitive skills in the same activities, lessons, or investigation(s).</li> <li>Lessons, activities, and investigation(s) in the unit, when taken together, represent multiple opportunities for students to construct meaning, apply content, and practice higher-order cognitive skills.</li> </ul>	3
2	<ul style="list-style-type: none"> <li>Lessons and activities require students to mainly rely on their teacher, text, or more capable peers to determine the accuracy of their results; the focus of authority for content knowledge is frequently focused on a source other than the students' reasoning.</li> <li>Lessons and activities require students to engage in inquiry and problem solving by following a predetermined strategy/format or that are not a cohesive part of the larger unit, which limits the opportunity for deep content learning.</li> <li>Activities/tasks/investigations are staged so that students are mostly passive participants following procedures or completing activities they know how to accomplish.</li> </ul>	<ul style="list-style-type: none"> <li>The unit is designed to target both content and higher-order cognitive skills, however, the lessons, activities, and investigations overemphasize content or cognitive skills at the expense of the other learning targets.</li> <li>Lessons, activities, and investigation(s) in the unit, when taken together, provide inconsistent opportunities for students to construct meaning, apply content, and practice higher-order cognitive skills.</li> </ul>	2
1	<ul style="list-style-type: none"> <li>Lessons and activities require students to rely on their teacher, text, or other authority to determine the accuracy of their results, observations, or conceptual understanding.</li> <li>Lessons and activities require students to engage in inquiry and problem solving that have no connection or fail to draw from content learning; inquiry and problem solving is treated as separate from the "regular" curriculum.</li> <li>Activities/tasks/investigations are staged so that students are passive participants, with rote memorization or mastery of vocabulary being the predominate goal.</li> </ul>	<ul style="list-style-type: none"> <li>The unit is designed to target only content or higher-order cognitive skills.</li> <li>Lessons, activities, and investigation(s) in the unit, when taken together, do not represent opportunities for students to construct meaning, apply content, and practice higher-order cognitive skills.</li> </ul>	1
0	Evidence either missing or too insufficient to score.	Evidence either missing or too insufficient to score.	0



Teacher Instructional Portfolio (TIP) Rubric

Score	Assessment for Learning	Score
4	<ul style="list-style-type: none"> <li>• Student conceptions and skills are frequently assessed prior to, during, and after lessons through intentional selection of formative and summative assessments that link clearly to learning goals; misconceptions or skill deficits are addressed through intentional teacher moves.</li> <li>• The unit is structured with frequent opportunities for students to self-assess, reflect on their observations, understanding, or teacher/peer feedback, and monitor their progress toward learning targets.</li> <li>• The unit's assessments consistently have appropriate cognitive demand, represent appropriate learning progressions, are aligned with designated learning goals and standards, and generate evidence of performance criteria for meeting those goals.</li> </ul>	4
3	<ul style="list-style-type: none"> <li>• Student conceptions are frequently assessed prior to, during, and after lessons through intentional selection of formative and summative assessments that link clearly to learning goals.</li> <li>• The unit is structured with periodic opportunities for students to self-assess, reflect on their observations, understanding, or teacher/peer feedback, and monitor their progress toward learning targets.</li> <li>• The unit's assessments primarily have developmentally appropriate cognitive demand, represent appropriate learning progressions, are aligned with designated learning goals and standards, and generate evidence of performance criteria for meeting those goals.</li> </ul>	3
2	<ul style="list-style-type: none"> <li>• Student conceptions sometimes are assessed prior to, during, and after lessons through intentional selection of formative and summative assessments that link to some, but not all learning goals.</li> <li>• The unit is structured with opportunities for students to self-assess, or monitor their progress toward learning targets, but these opportunities are infrequent or superficial.</li> <li>• The majority of the unit's assessment are aligned with designated learning goals and standards and generate evidence of performance criteria for meeting those goals, however, a few assessments are either not appropriate in terms of cognitive demand, not representative of learning progressions, or misaligned with goals/standards.</li> </ul>	2
1	<ul style="list-style-type: none"> <li>• Student conceptions are infrequently or not assessed prior to, during, and/or after lessons through intentional selection of formative and summative assessments.</li> <li>• The unit not is structured with opportunities for students to self-assess, or monitor their progress toward learning targets.</li> <li>• The unit's assessments primarily are either not aligned with designated learning goals and standards, not representative of learning progressions, fail to generate evidence of performance criteria for meeting those goals, or are not developmentally appropriate in terms of cognitive demand.</li> </ul>	1
0	• Evidence either missing or to insufficient to score.	0

## Appendix c

### Teacher Instructional Practices: Qualitative Coding Scheme

Code	Evaluation Criteria	Description
General Instruction Practices	Direct Instruction	Teacher provides direct instruction related to students' learning.
	Formative Assessment	Teacher formatively assesses students' learning.
	Scaffolding	Teacher provides scaffolding for students' learning.
Data Analysis Specific Instruction Practices	Direct Instruction (1)	Teacher provides direct instruction related to students' ability to correctly and thoroughly analyze their results, draw valid and comprehensive conclusions, and/or communicate their findings.
	Formative Assessment (2)	Teacher formative assesses students as related to students' ability to correctly and thoroughly analyze their results, draw valid and comprehensive conclusions, and/or communicate their findings.
	Scaffolding (3)	Teacher provides scaffolding related to students' a students' ability to correctly and thoroughly analyze their results, draw valid and comprehensive conclusions, and/or communicate their findings.
Data Analysis Best Practices	STEM Collaboration (4)	Teacher involves collaborative or integrated activities with other science, technology, engineering, or mathematics teachers or curriculum.
	Multiple Data Sources (5)	Teacher has students working with more than one data source (i.e. both first and second hand data).
	Multiple Data Types (6)	Teacher has student work with multiple data types (i.e. both covariant and non-covariant data)

(1) Based on the research of Keiler (2007) and Lovett and Chang (2007)

(2) Based on the research of Ruiz-Primo and Furtak (2007)

(3) Based on the research of Germann and Aram (1996)

(4) Based on the research of Keiler (2007)

(5) Based on the research of Hug and McNeill (2008)

(6) Based on the research of Kanari and Millar (2004)

Level	Description
0	No evidence of evaluation criteria
1	Teacher vaguely mentions code item
2	Teacher explicitly mentions evaluation criteria

## Appendix D

### From to 2011-2012 Official Scientific Inquiry Scoring Guide

#### High School Level

	<b>SI- Analyzing and Interpreting Results</b> <i>Summarize and analyze data, and identify uncertainties. Draw a valid conclusion, explain how it is supported by the evidence and communicate the findings of the scientific investigation.</i>
<b>5/6**</b>	<ul style="list-style-type: none"><li>• Draws a valid and comprehensive conclusion that addresses the question or hypothesis, identifies relationships in the data, and explicitly explains how the conclusion is supported by the data.</li><li>• Uses the results to analyze and critique the design and procedures providing significant sources of uncertainties and discuss how these might affect the results, and suggest insightful improvements, revisions or extensions.</li><li>• Communicates the findings using relevant terminology to report results, explain possible patterns within the data, and if needed justifies alternate reasonable explanations.</li></ul>
<b>4</b>	<ul style="list-style-type: none"><li>• Draws a valid conclusion that addresses the question or hypothesis and supports the conclusion explicitly using the data.</li><li>• Provides evidence that the design, procedures, and data have been reviewed to identify sources of uncertainties and discuss how these might affect the results.</li><li>• Communicates the findings using relevant terminology to report results, identify possible patterns within the data, and propose reasonable explanations.</li></ul>
<b>3</b>	<ul style="list-style-type: none"><li>• Draws a conclusion that addresses the question or hypothesis but is only partially supported by the evidence.</li><li>• Provides minimal evidence that the design, procedures, and data have been reviewed to identify sources of uncertainties.</li><li>• Communicates the findings using overly general terminology to report results and propose reasonable but incomplete explanations.</li></ul>
<b>1/2*</b>	<ul style="list-style-type: none"><li>• Draws a conclusion that is not clearly related to the question or hypothesis and is minimally supported by the evidence.</li><li>• Provides incorrect evidence that the design, procedures, data have been reviewed to identify uncertainties.</li><li>• Communicates the findings with inaccurate terminology to report results or proposes inaccurate explanations.</li></ul>

\*\*5 for preponderance (most) completed, 6 for all completed.

\* 2 for preponderance (most) completed, 1 for less completed or missing.

2011-12 December 2010

For scoring guide inquiries contact [Leslie.phillips@state.or.us](mailto:Leslie.phillips@state.or.us) or 503-947-5835

## Appendix E

### SWS Data Analysis: Qualitative Coding Scheme

Code*	Evaluation Criteria**	Description
Analyzing Results	Data Source	Student identified where the data came from, either by claiming ownership or by stating who collected the data
	Data Agreement	Student discussed agreement between data measurement and data discussion
Drawing Conclusions	Conclusions	Student drew conclusions based on the data
	Limitations of Data	Student discussed the limitations of the experiment
Communicating Results	Content Knowledge	Student related data discussions to content knowledge
	Patterns/Inferences	Student identified patterns or inferences from the data

\* Based on the codes of Germann and Aram (1996)

\*\*Based on the codes of Hug and McNeill (2008)

Level	Description
0	Student does not mention evaluation criteria
1	Student inaccurately or vaguely mentions evaluation criteria
2	Student accurately or explicitly mentions evaluation criteria